2019

DECISIONS LEADING TO INEQUALITIES IN THE JUVENILE JUSTICE SYSTEM: USING DATA TO PREDICT DELINQUENT OUTCOMES, INFORM DECISIONS, AND REDUCE DISPARITIES FOR JUSTICE-INVOLVED YOUTH

Patrick David McKay
The University of Montana

DECISIONS LEADING TO INEQUALITIES IN THE JUVENILE JUSTICE SYSTEM: USING DATA TO PREDICT DELINQUENT OUTCOMES, INFORM DECISIONS, AND REDUCE DISPARITIES FOR JUSTICE-INVOLVED YOUTH

By

Patrick David McKay

Master of Arts - Sociology, University of Montana, Missoula, MT, 2012

Bachelor of Arts - Sociology (Criminology), University of Montana, Missoula, MT, 2009

Dissertation

presented in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Interdisciplinary Studies

The University of Montana

Missoula, MT

May 2019

Approved by:

Scott Whittenburg,

Graduate School Dean

Dusten Hollist, Co-Chair

Sociology

Jackson Bunch, Co-Chair

Sociology

David Schuldberg

Psychology

Anna Klene

Geography

Brian Steele

Math

Bart Klika

Social Work

DEDICATION

This is for Professor Dusten Hollist. Thank you for years of guidance and friendship. You are truly missed.

ACKNOWLEDGEMENTS

I would like to acknowledge everyone who made the completion of my degree possible. First, thank you to my committee members: Dusten Hollist, Jackson Bunch, David Schuldberg, Anna Klene, Brian Steele, and Bart Klika. The fact that you blindly accepted to be on my committee speaks volumes about the quality of your character and your commitment to your students at the University of Montana. Thank you for your support on my committee and all that you do.

A special thanks to Jackson Bunch who pulled me across the finish line and was a good friend to me when the worst had happened.

I would also like to acknowledge and thank Dusten Hollist. Over the past 10 years you were my undergrad and graduate professor, thesis advisor, dissertation chair, colleague, mentor, and friend. You are the reason I enrolled in this program and have played a massive role in my life. Thank you for the time that you selflessly gave to me, I will be forever grateful.

I would also like to thank my entire family for all of your support. I'm lucky to be surrounded by such great people. I love you all.

Lastly, I would like to thank the person who motivates and supports me every single day, my wife Sydney. You are my inspiration in life and I could not be where I am without you. Thank you for always being there for me. I love you.

ABSTRACT

McKay, Patrick, Ph.D., Spring 2019 *Interdisciplinary Studies*

Decisions Leading to Inequalities in the Juvenile Justice System: Using Data to Predict Delinquent Outcomes, Inform Decisions, and Reduce Disparities for Justice-Involved Youth

Co-Chairperson: Dusten Hollist

Co-Chairperson: Jackson Bunch

This dissertation focuses on how decisions are made in the juvenile justice system and what strategies can be used to make more informed decisions in the future, in three separate investigations. (1) To begin, eight years of initial detention decisions (*N*=26,128) were collected to determine what factors affect the arresting officers' decision to detain youth at arrest. Findings reveal that the severity of current offense and prior offending history are the greatest predictors of initial detention. However, results demonstrate that race and the distance from a detention facility also influence these decisions. Non-white youth are more likely detained than their white counterpart holding relevant factors constant. Additionally, the closer an arrest takes place to a detention facility the more likely a youth is detained. This geographical variable is unique to this analysis and is found to mediate the effect that other geographic variables have on detention decisions (rural/urban classification). Also, evidence suggests that there is a greater disparity in the use of initial detention between white and non-white youth in areas that are closer in proximity to a detention facility. (2) Next, this investigation utilized six years of juvenile intakes (*N*=3,121) to create and validate a juvenile risk screener to predict recidivism in a one year period of risk. Items included in the risk screener were selected based on a series of bivariate and multivariate analyses. Out of the 246 risk factors eligible for the screener, seven factors were found to be important predictors for recidivism and were included in the risk screener. Validation measurements demonstrate comparable, and at times increased, prediction accuracy from the currently used, significantly larger risk assessment. (3) The final study analyzes an experimental method of weighting risk factors using conjunctive analysis of case configuration (CACC). This study demonstrates the importance of risk factor combinations and how these combinations affect the outcome. Using this weighting strategy, results found slight improvement in prediction accuracy over a simpler 0, 1 scoring method and a more complex score from logistic regression.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

The United States represents less than five percent of the global population but has close to one quarter (22%) of the world's prisoners (Walmsley, 2013). By the end of 2013, one in 35 adults in the U.S. were under supervision by the justice system (Glaze & Kaeble, 2014). Juvenile justice figures are just as staggering. According to the Annie E. Casey Foundation (2013), the U.S leads the industrialized world in the rate at which youth are detained at 225 per 100,000 juveniles. Criminological research has overwhelmingly focused on why juvenile offenders commit crime and the institutions that push and pull them into delinquency. Less work, however, has focused on the decisions made within the justice system after an offense has been committed, decisions that contribute to the above cited mass incarceration numbers. State and federal justice systems are a maze of decision points. Decisions made within the justice system begin at arrest and continue on to include such things as the determination of secure placement, service type, and supervision level, to name only a few. There is clearly a need to understand how these decisions are made, if decisions discriminate against certain groups of people, and what strategies can be deployed to curb over incarceration in the future.

Chapter 2 of this dissertation begins by detailing an investigation into the initial decision to detain a youth who becomes justice involved in Montana. Montana law enforcement officers have the discretion to detain youth before the initial court hearing if the officer believes the youth is a risk to public safety or a risk to themselves (MCA §41-5-102.3). There is a general assumption that initial detention decisions are based on legal factors such as offense severity and past criminal history. Study One analyzes initial detention decisions for 26,128 unduplicated youth to determine if decisions are objectively made based on current offense and criminal history, or if there are extralegal factors that contribute to this decision.

With the acknowledgment that disparities among certain groups of people are apparent at almost every stage of the justice system (see Sickmund, 2009), risk assessments are increasingly being called upon with the intention of creating a more structured decision-making approach (Miller & Maloney, 2013). Risk assessments—specifically recidivism risk assessments—play three roles in the justice system: (1) informing decisions regarding the detention of high-risk offenders, (2) informing decisions regarding the supervised released of low-risk offenders, and (3) informing decisions designed to reduce the risk posed by the offender (Singh, Kroner, Wormith, Desmarais, & Hamilton, 2018).

Chapter 3 focuses on the second role of recidivism risk assessments in the justice system. Utilizing data collected from 3,121 youth in Montana, Study Two creates and validates a risk assessment for Montana Youth Court Services juvenile probation departments to inform decisions regarding the supervised release of low-risk offenders. Currently, the Pre-screen Back On Track (pre-screen BOT) is used to predict recidivism risk. This assessment was created in the State of Washington and borrowed verbatim by Montana. Tool fidelity is demonstrated to be low and officers lack the necessary buy-in to make this assessment usable (McKay, Hollist, Bunch, Acton, Tillman, & Harris, 2015). The objective of the study detailed in Chapter 3 is to develop an assessment with significantly fewer risk factors than its predecessor and maintain, if not increase, recidivism risk prediction accuracy. The resulting product of Study Two is a quick assessment that is a valid predictor of recidivism capable of encouraging buy-in from officers in the field.

Since the creation of the first justice system risk assessment instrument (Burgess, 1928), researchers have examined ways to increase the prediction accuracy of assessments by locating additional factors that are more predictive of risk and ways to add meaningful weight to these risk factors. Contemporary literature has been successful at utilizing statistical techniques to locate factors that are the most predictive of risk; however, researchers have yet to find a superior way to add meaningful weight to risk factors that significantly outperform simple techniques.

The final study presented here (Chapter 4) investigates a new strategy to add weight to a risk assessment instrument using conjunctive analysis of case configuration (CACC; see Miethe, Hart, & Regoeczi, 2008). As detailed in the academic literature on this subject matter, CACC has never been used as a weighting technique, and it is unique from other weighting strategies. Unlike past techniques that try and add weight to individual risk factors, typically based on factor strength at predicting the outcome of interest (see Gottfredson & Gottfredson, 1980), the new strategy detailed here adds weight to combinations of risk factors that are shown to be the most predictive of recidivism. Chapter 4 utilizes the data (N= 3,121) and the newly created risk assessment detailed in Chapter 3 to examine this new technique.

Overall, research presented in the three studies discussed here demonstrates the process necessary to improve decision making within the justice system. This dissertation begins by demonstrating the need for objective informed decision making in the justice system (Chapter 2), continues by developing a data driven strategy to improve decisions (Chapter 3), and proceeds by researching and developing new ways to improve these strategies into the future (Chapter 4).

Chapter 2

Initial Detention Decisions in a Rural Frontier State

In Montana, the arresting officer has the discretion to detain youth at the point of arrest, before the youth's initial court hearing (MCA §41-5-322.2). This process is known as initial detention or pre-adjudicatory detention. Legally, youth can be detained if they are a danger to themselves or the public or they are believed to be a flight risk before their initial court hearing (MCA §41-5-102.3). There is a general assumption made by the public that these decisions are objective and based on legal factors such as the severity of arresting offense and the legal history of the youth. However, literature on initial detention demonstrates that other factors outside the scope of law also contributes to these decisions.

Race is one such factors found to influence decisions in the justice system (e.g., Bishop, 2005). Black youth are arrested for violent crimes at a higher frequency (Puzzanchera, 2009), more likely to be seen before juvenile court, formally processed, sent to residential placement, and sent to the criminal court compared to white youth whom are similarly situated (Sickmund, 2009). Moreover, a significant body of research has confirmed that minority youth, specifically black youth, are more likely to be detained than their white youth counterparts (Armstrong & Rodriguez, 2005; Bishop & Frazier, 1996; Bortner & Wornie, 1985; McGuire, 2002; Rodriguez, 2007; Wordes, Bynum, & Corley, 1994; Wu, 1997).

There is consensus that disproportionate minority contact (DMC) exists in certain decisions points in the juvenile justice systems (Ayres & Borowky, 2008; Fite, Wynn, & Pardini, 2009; Kakade, Duarte, Liu, Fuller, Drucker, Hoven, & Wu, 2012; Leiber, 2002). The cause of DMC, however, is still debated. On one side of this argument, the *differential offending perspective,* states that differences found in the justice system are due to the varied amounts and severity of crime committed by white and minority youth, respectively (see Development Service Group, 2014). This perspective examines individual, family, and environmental factors that may be contributing to delinquency. Alternatively, *the differential treatment perspective* argues that minority youth are simply treated differently (harsher) than white youth in the justice system (see Leiber, 2003; Pope & Feyerherm, 1990). The current study explores the differential treatment perspective in the juvenile justice system by examining initial detention decisions made for white and minority youth.

While less common in the literature, geography, like race, has been shown to impact decisions made in the justice system. Feld (1991), for example, argued urban juvenile justice systems engage in more formal social controls with increased levels of bureaucratization compared to their suburban and rural counterparts. Differing levels of social controls lead to inequalities between these geographic areas. Other studies have come to similar conclusions, demonstrating a divide between rural and urban areas as the causal factor of differences in justice system decisions (Blackmon, Robins, & Rhodes, 2016; Johnson & Secret, 1995).

If geography does influence detention decisions, one would expect these effects would be inflated in states where small populations are spread over large geographic regions (e.g., Montana). However, no research has examined the effect distances have on decisions in the justice system. Instead, geography has only been used in the literature to describe differences between rural and urban areas. This study fills the void in the literature by examining a spatial variable that has not been included in prior studies: the distance an officer must travel to securely detain a youth in initial detention. Distance to needed services (e.g., healthcare) is a fundamental barrier to receiving or providing services in states where population densities are low and geographical remoteness is high (see Bigbee, Gehrke, & Otterness, 2009). It is therefore likely, that distance will also impact decisions regarding services in the justice system.

Using data from eight years of juvenile intakes in Montana, this study exmaines the influence of race and geography on detention decisions. The current investigation adds to the understanding of geography's role in detention decisions by demonstrating the importance of distance and how distance mediates the effect that other geographical variables have on detention decisions. Additionally, this investigation adds to the understanding of DMC by demonstrating an interaction between race and geography in a rural state where American Indian youth make up over 50% of the minority sample.

Distance to Detention in Montana

Updates to the Montana Youth Court Act established a regional detention structure and funding mechanism for juvenile detention in Montana. The passage of this act transferred county control of probation to a state function organized into 22 districts courts. The 22 districts were then organized into five detention regions. Figure 1.1 presents the five regions along with the seven facilities used for youth initial detention in Montana.

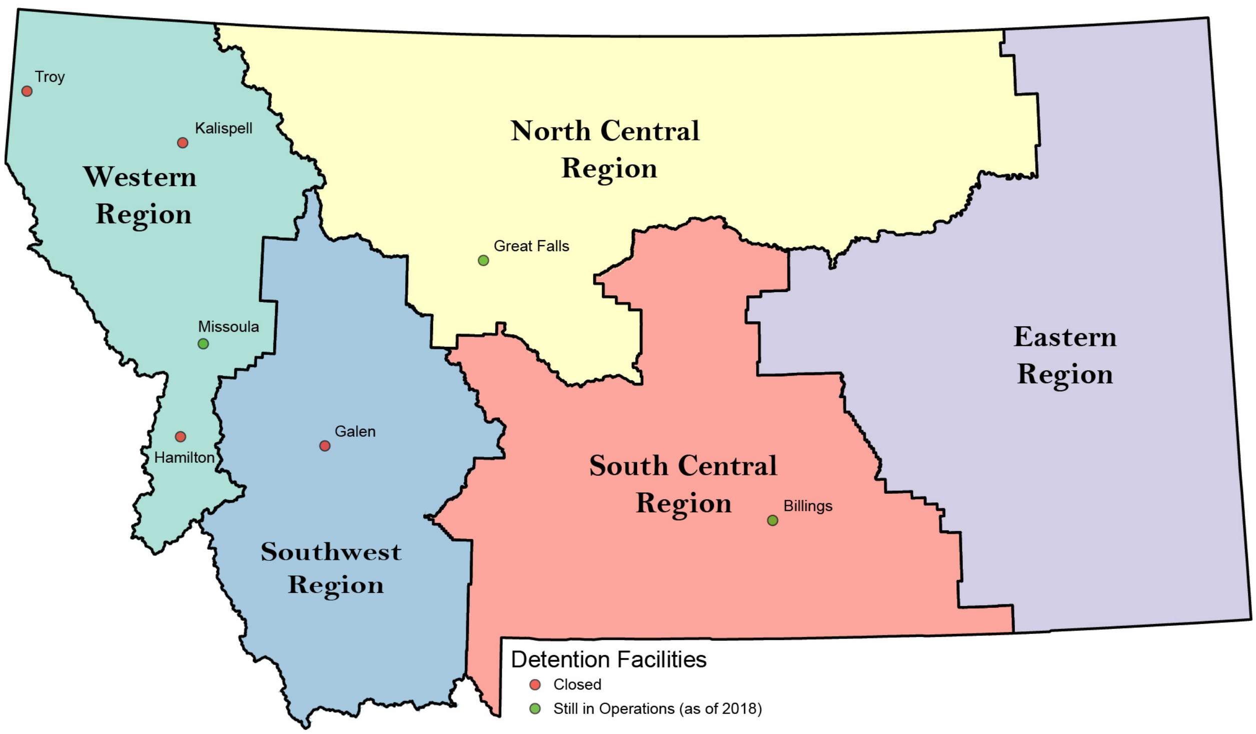

Figure 1.1: Map of Montana's regional detention structure and detention facility location from 2010 to 2018.

As of 2018, due to the high cost of running a detention facility and the decreasing use of youth detention, several facilities are no longer operational. In October 2011, the Hamilton detention facility closed, leaving the surrounding areas to use the Missoula facility. The Kalispell and Galen facilities both closed in January 2016. Areas that relied on Kalispell now use Great Falls or Missoula and communities that relied on the Galen facility now send youth to detention centers in Missoula, Great Falls, and

Billings. Additionally, in October of 2017 the Troy facility closed, leaving surrounding areas to send youth to the Missoula facility. In 2018, only three detention facilities in Montana remained: Missoula, Great Falls, and Billings.

In Montana, 45 of the 56 counties are classified as frontier.[1] A frontier area is one characterized by low-population density and high geographic remoteness (Cromartie, Nulph, Hart, & Dobis, 2012). Geographically, Montana is the fourth largest U.S. state measuring approximately 560 miles east to west and 320 miles north to south. Additionally, Montana is the eighth least populated state with approximately 1.05 million people (U.S. Census, 2017). Most of Montana's population are in the western portion of Montana with significantly smaller populations located in eastern Montana. The tremendous distance and low populations make it difficult for states like Montana to provide equal services to all residents.

A prime example of how distance affects Montana's juvenile justice system is evidenced by protocols undertaken by law enforcement in the small rural town of Havre in Hill County. Havre has a population of 16,463 people (U.S. Census, 2017). If a youth is arrested in Havre, and the arresting officer has determined that the youth must be detained until their court hearing, the arrested individual must be transported from the Hi-Line community adjacent to the Canadian border to the closest regional detention facility in Great Falls, which is 112.2 miles away. It then takes two hours to transport the youth one way, four hours for the officer to make the round trip, and an additional four hours for the officer to pick up the youth for their probable cause hearing the following day.[2] Time estimates cited here are based on good road conditions, which, in Montana, may be the case for only half of the year. The total time an officer

[1] Frontier counties are defined as those counties with a population density of fewer than seven people per square mile (see Cromartie, Nulph, Hart, & Dobis, 2012).

[2] It must be noted that certain counties across Montana use MetNet, a video conference system for youth to appear in court who are detained in facilities located further away. In instances where MetNet is utilized, some travel time may be reduced. The use of MetNet does not impact the initial distance an officer must travel to detain, or the distance required to bring the youth back home if no probable cause was found to hold them or they are otherwise released from detention.

from Havre would spend on transportation alone in such a scenario is approximately eight hours. Almost 40% of Montana's 56 counties (22) have a distance to detention that is equal to or exceeds the Havre scenario articulated here.

Montana youth are held in detention after an arrest if they are at risk of committing an additional offense, are a risk to themselves, or are a flight risk before their probable cause hearing. As the 2017 Montana Code Annotated states, youth should be separated from their parents by law enforcement "only when necessary for the welfare of the youth or for the safety and protection of the community" (MCA §41-5-102.3). In most cases, youth will be detained for less than 24 hours while awaiting the probable cause hearing. Time in detention can extend up to five days, however, in the case of weekends and holidays.

Montana laws cited above provide basic parameters for when and why a youth should be detained before a probable cause hearing. Problems arise, however, when statutory terminology allows for flexibility or is otherwise ambiguous. Language such as "only when necessary" or "believes on reasonable ground" is specifically troubling, because it lacks concrete definition. Individual police and or probation officers are therefore left to define these terms based on experience and training. Due to the flexibility in these definitions, personal biases that officers may not even be aware of (i.e. implicit bias) and situational contexts (e.g., distance to detention) may also influence the definitions of these law on an individual basis. What can occur, and what has been shown to occur in contemporary literature, is differential treatment of youth based on factors that should not contribute to these important decisions.

Initial Detention in the Juvenile Justice System

Due to the poor outcomes experienced by youth placed in initial detention, academic literature urges secure placement be utilized as a last resort only after all other options are exhausted and an objective standardized process has been followed (e.g., Mendel, 2009). One of the biggest problems with initial detention is that it increases the likelihood of an individual being treated more harshly at subsequent stages of the juvenile justice system (Bishop & Frazier, 1996; Holman & Ziedenberg, 2006; Mendel, 2009; Rodriguez, 2010). For example, youth held in initial detention are more likely than those who are not detained to be formally charged (Bishop & Frazier, 1988), adjudicated, and committed to out-of-home placement (Bortner & Wornie, 1985; Frazier & Bishop, 1985; McCarthy & Smith, 1986; McGuire, 2002).

Critics argue that the overreliance on and misuse of detention represents a failed juvenile justice system response. Among the evidence cited to support this view is research suggesting that detention exposes youth to crime (Mendel, 2011), interrupts the normal "aging-out" process (Golub, 1990), increases risk of self-harm (Parent, Leiter, Kennedy, Livens, Wentworth, & Wilcox, 1994), complicates mental health treatment (Forrest, Tambor, Riley, Ensminger, & Starfield, 2000; Sickmund, Sladky, Kang, & Puzzanchera, 2011), and is less effective and more expensive than detention alternatives (e.g., house arrest or electronic monitoring; Parent, Leiter, Kennedy, Livens, Wentworth, & Wilcox, 1994; Shelden, 1999).

One justification for detaining youth is the deterrence effect.[3] Current research supports the position that punishment and sanctions do not work as a deterrent for committing crime (Lipsey & Cullen 2007; Loughran, Mulvey, Schubert, Fagan, Piquero, & Losoya, 2009; Mackenzie, Wilson, & Kider, 2001) and can produce unintentional harmful effects (Gatti, Tremblay, & Vitaro, 2009). For example, several studies have shown that secure detention increases recidivism rates (Bezruki, Varana, & Hill, 1999; Dishion, McCord, & Poulin, 1999; Homan and Ziedenberg, 2006; Lipsey 2009; Mendel 2009; Vieira, Skilling, and Peterson-Badali, 2009), decreases school performance, and decreases the likelihood of success in later occupational endeavors, (Chung, Little, & Steinberg, 2005; Holman & Ziedenberg, 2006; Mendel, 2009). The over representation of youth with mental health disorders make these findings even more troubling.

[3] As stated in Montana Code Annotated, the goal is "To prevent and reduce youth delinquency through… immediate, consistent, enforceable, and avoidable consequences …" (MCA §41-5-102ab).

Between 65% and 70% of adolescents in the juvenile justice system have a mental health disorder (Cocozza & Shufelt, 2006; Cocozza & Skowyra, 2000; Teplin, Abram, McClelland, Dulcan, & Mericle, 2002; Wood, Foy, Goguen, Pynoos, & James, 2002). Meanwhile, in the general population, approximately 20% of adolescents have a diagnosable disorder, with only 5% to 9% with a serious impairment (Friedman, Katz-Levy, Manderscheild, & Sodheimer, 1996). Cocozza and Shufelt (2006) found in their study that 79% of adolescents found to have a mental health disorder in the juvenile justice system had two or more disorders. Additionally, isolation as a result of placement in detention can exacerbate the symptoms of mental illness and can provoke mental illness in youth with previously absent symptomology (Tandy, 2014).

Disproportionate Minority Contact

Literature on secure placement has consistently shown that it is disproportionately used against minority youth (Bishop & Frazier, 1996; Bortner & Wornie, 1985; Leiber, 2003; Leiber, Bishop, & Chamlin, 2011; Leiber & Fox, 2005; Rodriguez, 2010; Sickmund, Snyder, & Poe-Yamagata, 1997; Wu, 1997). In 2015, black youth were detained at one and three-quarter times the rate of white youth in the state of Maine and the broader category of non-white youth were detained at more than two and one-half times the rate of white youth. The differences could not be explained by age, gender, or offense class (King, Shaler, & Dumont, 2015). In addition, researchers have found that African-American youth were two times more likely to receive secure detention center placement than non-African-American youth even when a standardized risk assessment was taken into account (Mallet & Stoddard-Dare, 2010). Moreover, American Indian youth in Montana were more likely to be adjudicated delinquent and more likely to be securely confined when compared to their white counterparts (Hollist, Coolidge, Delano, Greenwood, King, McLean, McKay, Harris, Burfeind, & Doyle, 2012).

Similar to placement in secure detention, evidence of racial and ethnic disproportionalities are documented for initial detention decisions (Leiber & Boggess, 2012). For example, Rodriguez (2010) found that Latinos (1.23 times), American Indian (1.93 times) and black youth (1.49 times) were more likely than their white counterparts to be held in initial detention while holding several factors related to detention constant. Because initial detention is used disproportionality against minority youth, it is believed that there is a compounding effect where racial disproportionality is even more extensive in later stages of the justice system (Pope & Feyerhem, 1993; Sampson & Lauritsen, 1997). Studies have demonstrated that minority youth historically received harsher dispositional sanctions when compared to white youth (Hockenberry & Puzzanchera, 2015; Moor & Padavic, 2010) and these subsequent sanctions (e.g., initial detention) increase recidivism and other poor outcomes (Aizer & Doyle, 2013; Campbell, Barns, Mandalari, Onifade, Campbell, Anderson, Kashy, & Davidson 2018; Lambie and Randell, 2013; Munyo, 2015).

Some research has contended that disproportionate minority contact (DMC) is inconsistent across the various stages of the juvenile justice system, in particular at adjudication and disposition (Dejong & Jackson, 1998; Leiber, Bishop & Chamlin, 2011; Leiber & Fox, 2005; Leiber & Johnson, 2008; Moore & Padavic, 2010). However, DMC is consistently found at the arrest point of contact where initial detention decisions are made (Ayres & Borowky, 2008; Fite, Wynn, & Pardini, 2009; Huizinga et al., 2007; Kakade et al., 2012; Leiber, 2002).

Research on DMC has moved beyond the question of whether it is occurring to the question of why it is occurring (see Bishop, 2005). Often the reason DMC occurs is explained through two perspectives: (1) *differential offending;* or (2) *differential treatment*. "The differential offending perspective requires that causes of differential involvement be sought outside the court system by looking at the individual, family, and neighborhood factors that are related to offending" (Development Service Group, OJJDP, 2014 p. 3). Alternatively, the differential treatment perspective focuses on the structure of justice decision-making acts that are disadvantageous to minority youth (Leiber, 2003; Pope & Feyerherm, 1990). This perspective argues that differential treatment will still exist even after accounting for individual factors found in differential offending.

Gender

Gender has also received research attention in differential treatment of youth in the justice system (Gaarder, Rodriguez, & Zatz, 2004; Leiber & Peck, 2012; MacDonald & Chesney-Lind, 2001). Findings in that arena have been mixed, however, with evidence showing both males and females receiving more severe treatment respectively in different stages of the criminal justice system (Chesney-Lind, 1977). An example of this is demonstrated in the research of Tam, Abrams, Freisthler, and Ryan (2016) where females were more likely to be sentenced to group homes compared to males for certain dispositions and males were more likely to be sentenced to correctional facilities compared to females. Additional research has found no difference between the treatments of gender in the juvenile justice system (Dannefer & Schutt, 1982; Kempf-Leonard & Sontheimer, 1995).

Geographical Disparities in the Justice System

Race and gender are common in research examining differential treatment in the juvenile justice system. Far less attention has been given to the impact of geographic location. In 1991, Barry Feld coined the term "Justice by Geography" in which he argued that that disparities in juvenile justice decision making emerged in court, based on different geographical locations. Feld found in Minnesota that initial detention was used twice as often in urban counties as in suburban and rural ones. Feld believed that these differences were explained by the increased level of bureaucratization that urban justice systems exercised, which, in turn, increased formal social control more so than their suburban and rural counterparts. Johnson and Secret (1995) found rural county court judges with general jurisdiction were more punitive in their sanctions than specialized urban county juvenile courts. Similarly, Blackmon, Robison, and Rhodes (2016) found that rural youth were also at an increased risk of encountering the juvenile justice system compared to urban youth.

Several studies have found that race and geographic location together play an important role in juvenile justice system decisions. Shook and Goodkind (2009) describe is an interaction between race and geography. They found that black and white youth from urban areas present similar initial detention rates.

However, white youth from suburban areas are far less likely to be detained than their urban white counterparts. And suburban black youth are just as likely to be detained as their urban black counterparts. Leiber, Richetelli, and Feyherm (2009) posit that geography impacts decisions on minority youth because minority youth tend to live in areas that have stricter law enforcement or harsher judges, compared with jurisdictions where white youth live. Similarly, Kauffmann (1997) found in Massachusetts that police tend to patrol urban minority neighborhoods more aggressively than suburban areas where fewer minorities reside. Gordon (2016) also found disparities based upon race and geographic locations and argued that these differences were due to a lack of statewide standards which led to inconsistent responses across the state of Utah.

Current Research

Prior research has demonstrated that race and geography impact detention decision in higher populated states and cities. The geographic variable most commonly analyzed in prior work is the boundaries separating space into rural, suburban, and urban areas. However, these findings cannot be generalized to less dense rural states where small populations are spread over large geographic regions. In addition, prior research has focused mainly on differences between white and African American populations with few that have analyzed populations where American Indians are the largest minority race. These factors likely have consequences when it comes to decisions in the juvenile justice system and for this reason, the current analysis presents vital insights into law enforcement decision making that will fill gaps in the current literature. This study examines the question "to what extent do race and geography impact detention decisions in less dense rural states." The following hypotheses will be examined in the current research:

> *Hypothesis 1:* The distance an officer must travel to detain a youth in initial detention will decrease the likelihood the youth will be detained.
>
> *Hypothesis 2:* The relationship between rural/urban areas and detention will be mediated by the distance an officer must travel to detain a youth.
>
> *Hypothesis 3:* Race affects detention decisions where non-white youth will be detained at a higher rate than white youth.

Hypothesis 4: Race has a moderating effect on distance to detention where distance influences detention decisions differently for white and non-white youth.

Methods

Sample

All youth citations for an offense between January 1, 2010 through December 31, 2017 are included in the original dataset (*N*=49,854). Approximately half (52.4%; *n*=26,128) of the youth during this time period had multiple violations or were "duplicates." A duplicate is the term used here when more than one intake related to one youth occurred during the eight-year timeframe. To ensure that no youth is included in the sample more than once, one case involving that individual is randomly selected and all other cases involving them are removed from the sample.[4] The final sample contains 23,726 unique youth intakes. Most data are queried from the Juvenile Court Assessment and Tracking System (JCATS), which is the juvenile justice data repository for Montana. The remaining data are collected from information gathered by the Montana Board of Crime Control and Montana State Library Geographic Information Clearinghouse. These data sources are described below.

Data

Independent variables queried for this analysis fall into either legal or extralegal factors. Legal factors are variables that measure current and past crime events. Extralegal factors are all other factors that are not within the scope of the law. The legal factors collected on youth in this investigation are the most serious offense alleged at intake and the history of prior intakes. Extralegal factors queried from the JCATS include location of intake (city/town, county, zip code), intake date, age at offense, gender, and race/ethnicity. The dependent variable in this investigation is whether the youth was detained at arrest, otherwise referred to as initial detention.

[4] A random selection of duplicate cases is used to remove any systematic error that could be included if only the first offense, most recent offense, or the most serious offense were to be included in the analysis.

The Montana Board of Crime Control provided the list of police and sheriff department locations and the regional detention facility that each department utilized for initial detention. This information is used to create the extralegal geographic factor "distance to detention." This variable measured how many driving miles a youth is away from a detention facility.[5] Using definitions from the Office of Management and Budget, Standards for Delineating Metropolitan and Micropolitan Statistical Areas (OMB, 2015), each case is given an extralegal rural/urban county variable that describes if the youth's intake occurred in a "non-core" county, a "micropolitan" county or a "metropolitan" county. Metro counties in Montana are the closest geographic area to "urban" counties and micro and non-core counties are commonly classified as "rural."

GIS data used to create all maps in this study were obtained from the Montana State Library Geographic Information Clearinghouse. These data include state and county boundary shape files and city/town point data.

Measures

For the following analysis the dependent variable of initial detention is measured dichotomously, where 1 = detained. Race is also represented by one dichotomous variable and coded such that white = 1, non-white = 0. Race is dichotomized to account for the small sample size of minority youth. The variable distance to detention facility is dichotomously coded[6] where 1 = youth is arrested outside 50 miles to a regional detention facility, and 0 = youth is arrested within 50 miles from a regional detention facility. To account for differential treatment by population type, three rural/urban variables are dichotomously coded: non-core = 1, else = 0; micropolitan (micro) = 1, else = 0; and metropolitan (metro) = 1, else = 0. Gender is dichotomously coded where 1 = male, 0 = female. Age at intake and intake year are coded as continuous variables.

[5] As detention facilities closed during the study period, the distance to detention variable was updated to reflect the new distance needed to travel.

[6] Distance to detention facility was originally coded as continuous variable but preliminary analyses demonstrate a natural break in detention rates between 0 to 50 miles and 51+ miles to a detention facility.

Characteristics of the youth's current offense are included to account for type and severity of offense. Felony is dichotomously coded where 1 = current offense is a felony, 0 = else; misdemeanor is dichotomously coded where 1 = current offense is misdemeanor, 0 = else; and status and technical offenses are coded into one dichotomous variable where 1 = current offense is a status or technical violation, 0 = else. Prior offending history is also accounted for as a dichotomous variable where 1 = youth has prior intake(s), 0 = youth does not have a prior intake.

Table 1.1 provides the descriptive statistics for this sample. The significant overrepresentation of non-white youth is noteworthy. Non-white residents represent approximately 10.9% of the population (US Census, 2017) in Montana but comprise 18.7% of the overall sample of juvenile justice citations between 2010 and 2017. Non-white youth in the juvenile justice population are predominately American Indian (68.8%), followed by Hispanic/Latino (15.9%), African American (11.5%), and other (3.9%).

Table 1.1: Demographic Characteristics (*N*=23,726)

Demgraphic Variable	**Category**	**Min**	**Max**	**Mean(st.dev)**
Age		5	17	15.01 (1.9)
		f	**%**	
Gender	Female	8,895	37.5%	
	Male	14,831	62.5%	
Race/Ethnicity	White	19,298	81.3%	
	Non-White	4,428	18.7%	
Intake Year	2010	4,023	17.0%	
	2011	3,353	14.1%	
	2012	3,127	13.2%	
	2013	2,550	10.7%	
	2014	2,453	10.3%	
	2015	2,448	10.3%	
	2016	2,826	11.9%	
	2017	2,946	12.4%	
Current Offense	Status/Technical	5,355	22.6%	
	Misdemeanor	16,045	67.7%	
	Felony	2,284	9.6%	
Designated Area	Non-Core	8,340	35.2%	
	Micropolitan	6,840	28.8%	
	Metropolitan	8,546	36.0%	
Distance to Detention Facility	Within 50 Miles	13,866	58.6%	
	Beyond 50 Miles	9,784	41.4%	
Detained	Detained	2,243	9.5%	
	Released	21,460	90.4%	

Analytic Strategy

The analysis begins by examining Montana's geography with the use of a choropleth map.[7] Figure 1.2 presents the choropleth map depicting the relationship between distance to a detention facility, rural/urban designation (metro, micro, non-core), and detention rates across Montana's 56 counties. Cartographic analyses portray a complex relationship between these three variables which the following analyses attempt to untangle.

Next, bivariate correlations, using Pearson's r, are examined between the independent variables (legal and extralegal factors) and the dependent variable (detention). The correlation of all variables in this analysis to the dependent variable and the significant number of independent variables correlated with each other justify the subsequent multivariate analysis examining the mediating effect of distance to detention on the relationship between rural/urban designation and detention decisions. This analysis continues by examining the moderating effect that race has on the relationship between distance to a detention facility and detention decisions (Baron & Kenny, 1986).

Finally, logistic regression is used in this analysis due to the dichotomous nature of the dependent variable (detained/not detained). Three logistic models are presented in Table 1.2. The first model includes all independent variables of interest except for the variable measuring the distance to a detention facility. This model was run to partially replicate past studies that found that rural/urban designated areas are significant predictors of detention. Model 2 incorporates all variables from Model 1 and includes the variable measuring distance to a detention facility. These results demonstrate how distance mediates the relationship between rural/urban designation and detention rates. Finally, Model 3 includes the interaction between race and distance to detention to explore how race moderates the effect that distance to detention has on detention decisions.

[7] Jenks natural breaks are used in ArcGIS to create detention rate categories.

Results

Figure 1.2 is a choropleth map of detention rate by county. White markers represent the regional detention facilities that were in operation during the 2010 through 2017 timeframe (see Figure 1.1 for city labels). Black markers represent sheriff and police departments across Montana. The line connecting these symbols are the straight-line distance between the departments and the regional detention facilities.[8] Each county is symbolized in color based on the detention rate calculated over the eight years of this study period where darker shades represent higher detention rates. Counties are also categorized as either metropolitan (red border), micropolitan (blue border), or non-core (grey border). Some counties have few intakes during the eight-year period making detention rate an unstable measurement across the state. [9] Eastern Montana counties with significantly smaller populations are at higher risk of this unstable measurement and caution should be used when examining the variation in detention rate on this map.

There are five counties defined as micropolitan (micro) and five counties defined as metropolitan (metro) with the remaining 46 counties defined as non-core. It is important to note that three of the seven detention facilities are located within metro counties, one facility is in a micro county, and two are in non-core counties.

As displayed in Figure 1.2, there does appear to be some clustering of higher detention rates surrounding each of the regional detention facilities. This is expected if distance to a detention facility influences detention decisions. This relationship is made more complex by the visual clustering surrounding micro counties. The best examples of clustering can be seen surrounding the Billings, Great Falls, and Missoula regional detention facilities which also happen to be in metro counties. Taken together it appears that that geography has an impact on detention rate across the state, but further

[8] Lines displayed in map reflect the distances when all facilities were in operation in 2010.
[9] For example, Carter County has only four intakes during the investigation period and none of these four youth were detained. This gives Carter County a detention rate of 0%. Alternatively, Petroleum County also has four intakes, with one of these resulting in detention of the youth. This gives Petroleum County a 25% detention rate.

analyses help explain the relationship between rural/urban classification (micro, metro, non-core) and distance to detention.

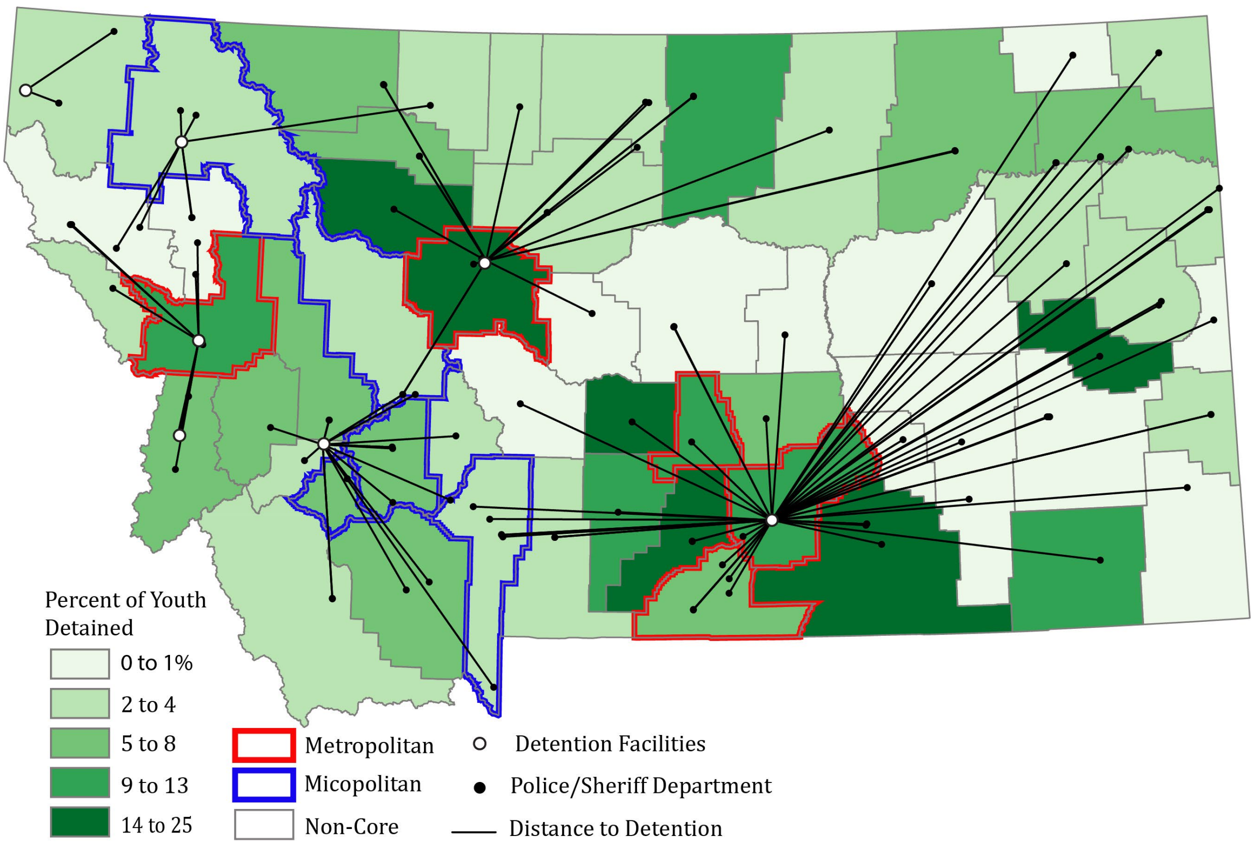

Figure 1.2: Map of the percent of arrested youth who were detained, the county designation, and distance to a detention facility in Montana 2010-2017.

Bivariate correlations are displayed for the variables of interest in Table A (Appendix A). The large sample size (n = 23,595) is a contributing factor to explain why each of the independent variables is significantly correlated with the dependent variable (detention). Being arrested in a metro county and outside of 50 miles to a detention facility are both shown to have the strongest correlations to detention for extralegal factors. Youth arrested for a felony offense demonstrate the strongest bivariate correlation to detention overall. Multivariate analysis is run to further explore the relationship between these factors and detention decisions.

Table 1.2 displays the results of binary logistic regression analyses of legal and extralegal factors on detention. [10] To be able to compare results found in the current research to prior studies, Model 1 only includes current offense, prior offense, demographics, and county designation (metro, non-core, and micro). Model 2 replicates Model 1 with the addition of the variable measuring distance to detention. Finally, Model 3 includes an interaction variable between distance to detention and race (white).

Legal factors are the strongest predictors of initial detention in both Models 1 and 2. It is apparent that officers consider the severity of current offense and the youth's prior intakes when making the decision to detain a youth prior to adjudication, as is the general assumption made by the public. A current felony offense is the greatest predictor of detention. Youth arrested for a felony offense, are 4.63 (Model 1) to 4.72 (Model 2) times more likely to be detained than youth arrested for a misdemeanor while holding all other factors in the model constant. Prior intakes are the second strongest predictor of initial detention. Prior intake(s) increase the likelihood of detention by 2.76 (Model 1) to 2.87 (Model 2) times. Alternatively, youth arrested for a status or technical violation are 53% (Model 1) to 57% (Model 2) less likely to be detained compared to youth arrested for a misdemeanor offense.

According to Model 1, being arrested in a metro county is the strongest extralegal predictor of initial detention. Youth arrested in a metro county are 61% more likely to be detained relative to a non-core county. Interestingly, youth from micro counties are 34% less likely to be detained relative to living in a non-core county. These results seem to be inconsistent with the logic that population size, based on these geographic areas, impacts detention decisions. Micro counties have higher populations than non-core counties but lower likelihood of detention than non-core counties in this model.

[10] Assumptions for logistic regression were investigated before proceeding with this analysis. This included collinearity diagnostics (Tolerance and VIF scores). Based on the recommended thresholds from Midi, Sarkar, and Rana (2010) all tolerance scores are below 0.2 and all VIF scores are below 2.5. This provides evidence that multi-collinearity is not a concern among any of the independent variables used in this analysis and logistic regression can proceed.

Table 1.2: Logistic Regression Analyses on Detention (*N* = 23,595)

	Model 1 Odds Ratio	Model 2 Odds Ratio	Model 3 Odds Ratio
Independent Variables			
Current Offense Felony	4.63***	4.72***	4.72***
Current Offense Status/Technical	.47***	.43***	.43***
Prior Intake	2.76***	2.87***	2.87***
Year of Intake	0.99	1.02	1.02
Age at Intake	1.13***	1.13***	1.13***
Male	1.09	1.10	1.10
Metro	1.61***	.96	.97
Non-Metro Micropolitan	.65***	.61***	.62***
White	.71***	.70***	.76***
Detention Facility Outside 50 Miles	--	.42***	.54***
Interactions			
White x Detention Outside 50 Miles	--	--	.72**

* $p \leq .05$; ** $p \leq .01$; *** $p \leq .001$

Certain demographic variables are also shown to be significant predictors of detention. According to both Model 1 and 2, white youth are approximately 30% less likely to be detained relative to non-white youth even when controlling for offense severity and offence history. Age is also found to be a significant predictor of initial detention. For each year increase in age there is a 13% increase in the likelihood of being detained. Gender does not impact the likelihood of being detained.

As stated above, Model 2 incorporates the geographic variable that measures the distance a detention facility is located from the place of arrest. The results from Model 2 suggest that when added to the model, distance to detention completely mediates the effect that being arrested in a metro county has on the likelihood of detention, relative to being arrested in a non-core county. The distance from a detention facility is shown to be the strongest extralegal factor on detention decision. Youth who are arrested further than 50 miles from a regional detention facility are 58% less likely to be detained relative to youth arrested within 50 miles while holding all other variables in the model constant.

Model 3 explores the moderating effect that race has on geographic location with the inclusion of the interaction between distance to detention and race (white). According to this model, there is a statistically significant interaction between these two variables. This suggests that while distance to detention affects the likelihood of being detained it does not affect all races equally. To illustrate this interaction, Figure 1.3 displays the estimated probabilities of initial detention inside and outside of 50 miles to detention facility, for a 17-year old male, arrested for a felony offense, with prior intakes, located in a metro county. The solid line presents the estimated probabilities for white youth and the dashed line presents the estimated probabilities for non-white youth.

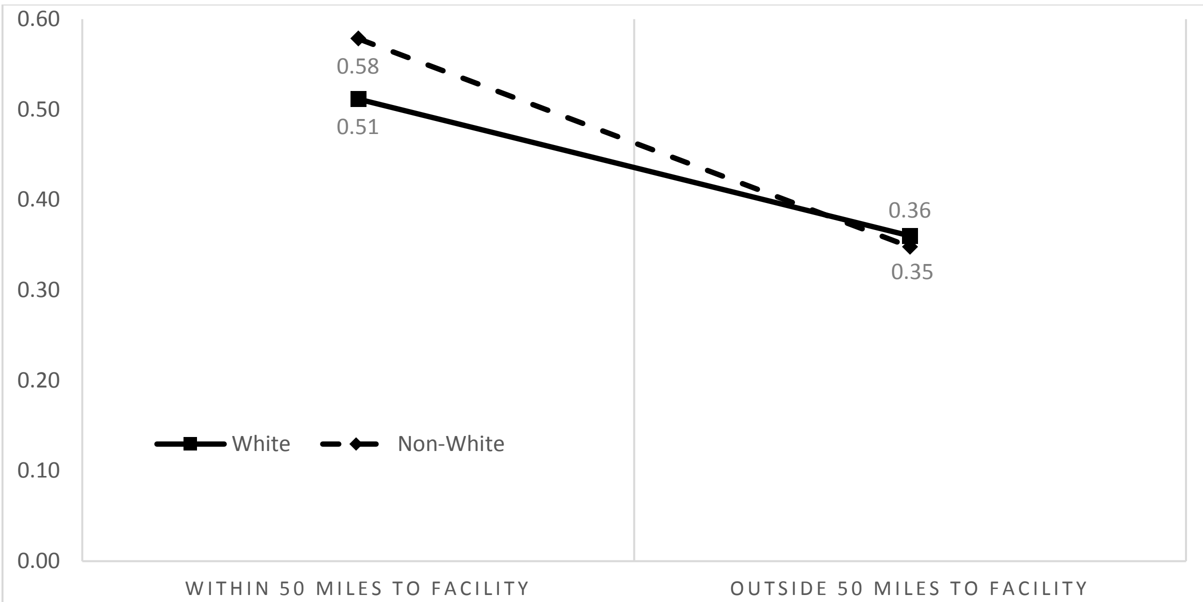

Figure 1.3: Graph of estimated probabilities of detention for a 17-year-old male, arrested for a felony offense, with prior intake(s), located in a metro county within and further than 50 miles to a detention facility for whites and not-whites.

As depicted in Figure 1.3, white and non-white youth share similar estimated probabilities when arrested outside of 50 miles to a detention facility (0.36 and 0.35 respectively). These probabilities increase dramatically for all youth when the arrest location is within 50 miles to detention facility. However, the increase in estimated probabilities of detention is greater for non-white youth. The result is a disparity, where non-white youth are more likely detained than white youth when they are within 50 miles to a detention facility while holding all other relevant factors constant.

Discussion

The decision of whether to detain a youth is the first of many made by juvenile justice system actors. Indeed, due to the impact detention has on subsequent outcomes, it may be the most important decision. As such, the decision must be objective and standardized to ensure that no youth is detained unnecessarily or treated unfairly relative to others who are similarly situated. Several studies have demonstrated a relationship between initial detention and severity of subsequent processing in the juvenile justice system (Bishop & Frazier, 1988, 1995; Bortner & Wornie, 1985; Frazier & Bishop, 1985; McCarthy & Smith, 1986; McGuire, 2002). When disparities exist at this initial stage, they are perpetuated in later processes of the juvenile justice system. The overall goal of this investigation was to determine how detention decisions are influenced in smaller rural state. The more that is understood about detention decisions the more than can be done to provide guidance in creating a more objective and fair decision process.

The results generally supported the hypotheses posed in this investigation. Geography and race are both important predictor of detention in a rural state. Contrary to prior studies that contend that a rural/urban divide is the cause for difference in detention decision, the current investigation demonstrated that distance to a detention facility explains both differences in detention decisions and the effect that rural/urban counties have on detention decisions. Geography must be examined in a different context when populations are small and less dense. In addition, race alone was found to be a significant predictor of detention but also moderates the effect that distance to detention facility has on detention decisions. These findings provide insight where racial disparities are most pronounced.

Geography's Impact on Detention Decisions

Prior research contends that geography does influence detention decisions (Feld, 1991; Johnson & Secret, 1995; Shook & Goodkind, 2009; Gordon, 2016). Lack of statewide standards (Gordon, 2016) and increased and decreased bureaucratization between urban, suburban, and rural populations (e.g., John & Secret, 1995; Feld, 1991) explain how geography can cause differential treatment in the justice system. This explanation is insufficient, however, when exploring the influences of detention decisions in a state comprised of rural counties.

Results discussed above provide evidence to support two hypotheses centered on geography's impact on detention decisions. First, the distance an officer must travel to securely detain a youth was found to be the strongest extralegal predictor of initial detention. The distance to a detention facility is considered when a detention decision is made and acts as a barrier to this decision when the distance is greater. The impact that proximity to a detention facility has on detention decisions pinpoints the troubling subjectivity of these decisions. Second, the distance an officer must travel to detain a youth explains the effect that living in a metro county has on detention decisions. Metro counties tend to be closer to or have their own detention facility located within them. When distance is considered, the effect that being arrested in a metro county has on detention is completely mediated. These findings are unlike prior studies that suggest differences in detention decisions are due to inherent difference within these rural and urban areas (e.g., Feld, 1991).

Although the current analyses provide limited evidence to explain the specific reasons for these results, the following two scenarios describes a potential explanation as to why geography is impacting detention decisions: (1) arresting officers, who are further away from a detention facility, are cognizant that the cost and time of detaining youth is high. The high costs provide a disincentive which encourages officers, located further from a facility, to either have a higher risk threshold or to spend more time determining if a youth is a serious risk to themselves or others. The increase in time may provide a more accurate picture of the youths risk and thus officers are able to rely less on initial detention; and/or (2) arresting officers closer to a detention facility do not have distance as a mitigating factor and rely more simply on basic indicators of risk (e.g., severity of offense and history of offending) to a greater extent. Further work is necessary to determine if one or both of these scenarios are adding to the different rates of detention between facilities closer and facilities further away.

The Impact of Race on Detention Decisions

Literature addressing disproportionate minority contact (DMC) demonstrates that detention and other processing decisions in the justice system are not fully objective or standardized and race has an impact on these important decisions (see Bishop, 2005; Davis & Sorensen, 2013; Kempf-Leonard, 2007; Leiber, Bishop, & Chamlin, 2011). One explanation for these disproportionalities are due to differential treatment between white and non-white youth in the justice system. This perspective holds that in the justice system minority youth are treated more severely than white youth based on their race or ethnicity alone (see Leiber, 2003; Pope & Feyerherm, 1990).

The findings from this research provide evidence to support the hypothesis that race affects detention decisions. A simple examination of the demographics in this sample show that non-white youth, largely American Indian youth, are significantly overrepresented in Montana's Juvenile Justice System. Moreover, once arrested, non-white youth are then at higher risk of being placed into initial detention relative to other white youth similarly situated.

These findings partially support the differential treatment perspective. Minority youth are at an increased risk of detention even when relevant factors (e.g., offense severity, prior offenses) are held constant. However, as described in prior work (Hollist et al., 2012) American Indian youth have other contributing factors that lead to increased rates of detention. One such example is an officer being unable to locate or contact American Indian parents or guardians at a higher rate. Family factors such as this fall into the differential offending perspective of DMC. While DMC is apparent in Montana, more work must be done to determine why this is occurring.

Interaction Between Race and Geography

Distances to detention acts as a protective factor for all youth when counties are further away but as discovered in this analysis, distance does not affect all youth equally. These findings support the final hypothesis that race moderates the effect that distance to detention has on detention decisions. When arrested further away from a detention facility, white and non-white youth are treated equally regarding detention decisions. As the distance to detention decreases, all youth are at an increased risk of detention, but non-white youth are at an even greater risk than white youth.

The explanation for the interaction between race and distance to detention is complicated. If the further a detention facility is located causes officers to have a higher risk threshold and/or spend more time investigating the underlying risk that the youth poses, then they may gain a greater objective understanding of the youth's actual risk. This deeper understanding may lessen the effect of implicit and overt biases that could be contributing to increased detention decisions based on the youth's race. Meanwhile, if less time is spent investigating risk for those youth closer to a detention facility it may be easier to associate the youth's race with an increased or decreased risk. This was the case for Bridges and Steen (1998) who found black youth were perceived as more aggressive and less cooperative than white youth, which lead to increased perception of risk.

Limitations and Future Research

This investigation's main limitation centered on the availability of data. Only certain factors are collected in the JCATS system on youth cited for an offense. This led to the potential omission of important factors impacting detention decisions. Additionally, some variables included in this analysis were limited in their definition. The most serious offense at intake was included but was broken down into three broad categories (status/technical offense, misdemeanor, and felony). Researchers did not have the ability to identify more specifically the nature of the offense. For example, Shook and Goodkind (2009) found youth who were arrested on a weapons charge was a significant predictor of initial detention. Similarly, the current investigation used the dichotomously coded variables of "prior intakes" as a measure of the youth's criminal history. There is likely a difference between youth who had one prior intake versus a youth who had multiple or more severe prior intakes. These differences were unable to be accounted for in the available data. While these limitations may help provided a more detailed picture of detention decisions, their inclusion would not likely change the overall conclusions drawn in this investigation.

Future research on initial detention should account for the limitations described above to determine if additional factors or factors defined in a narrower scope can further account for differences in detentions decisions. To gain a greater substantive answer as to why distance to detention facilities and race impact detention decisions, qualitative methods should be included in future investigations to provide context for the findings presented here. Finally, future research should focus on juvenile detention policy in rural frontier states like Montana to determine what strategies would reduce racial and geographical disparities.

The lost time and cost of transporting a youth to a detention facility in rural frontier states provides a disincentive to detention. Officers closer to a detention center do not have this disincentive and, if anything, it may be easier to detain, especially for youth who present a questionable risk level, a more severe current offense, or a longer history of offending. Such disincentives present significant questions about equity in the justice system.

Considering these lingering questions about equity among justice-involved youth, the following two areas of inquiry are recommended to inform juvenile justice policy:

- Investigate how detention decisions are made in areas further away from detention centers and standardize this process. A standardized process for determining risk and detention requirements based on areas already successful in this process could reduce overall detention rates across the state and reduce disparities between white and non-white youth.
- Create incentives to keep youth out of detention and disincentives to incarceration. Examples of such protocols could be a reward system for districts with lower detention rates, an increased cost to departments with high detention rates, or a detention limit that cannot be breached without triggering an internal investigation. Providing incentives and/or disincentive could reduce the present geographic disparities and further reduce the overall detention rate.

Conclusion

The overall goal of this investigation was to determine how detention decisions are influenced in the justice system in smaller rural states. The findings indicate detention decisions are largely influenced by legal factors such as current offense and history of offenses, as is expected. However, this investigation presented strong evidence to support the hypothesis that proximately to detention facility affects the likelihood of detention. Not only was distance to detention the greatest extralegal predictor of initial detention, this factor mediated the effect that other geographic variables had on detention decisions (i.e. arrested in metro county) which were found to be important predictors in prior studies. Moreover, as found in several studies, race was a significant predictor of detention. When holding current offense, prior offenses, and other important variables constant, non-white youth are still at an increased risk of detention compared to their white counterparts. Race moderated the effect that distance to detention has on detention decisions showing that there is greater disparity between white and non-white youth in areas that are closer to a detention facility than those areas further away. Equity for geographical or racial differences should be a priority for the juvenile justice system moving forward. Overall, this inquiry indicates that certain youth are at a greater risk of detention than other similarly situated youth based on factors that should not be used to make detention decisions. The juvenile justice system should strive to make these important decisions fair and just for all, because this decision will have a lasting impact on those youth detained.

Appendix A:

Table A: Bivariate Correlations of Detention, Predictors, and Control Variables (*N* =23,595)

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Detained at intake	1												
2. Felony	.21**	1											
3. Misdemeanor	-.05**	-.48**	1										
4. Status/Technical Offense	-.10**	-.18**	-.78**	1									
5. Year of Intake	-.02**	.00	-.02*	.02*	1								
6. Age at Offense	.07**	-.02**	-.04**	.06**	-.11**	1							
7. Male	.04**	.13**	.03**	-.12**	.00	-.03**	1						
8. Prior Intakes	.16**	.02**	-.05**	.04**	-.11**	.19**	.05**	1					
9. Detention Outside 50 Miles	-.12**	.02*	-.07**	.07**	.13**	-.00	.03**	-.02**	1				
10. Metro	.10**	-.03**	.13**	-.12**	.01	-.01	-.03**	.03**	-.62**	1			
11. Non-Metro Micropolitan	-.08**	-.01	.03**	-.03**	-.02**	.04**	.00	.02*	.23**	-.48**	1		
12. Non-Metro Non-core	-.03**	.04**	-.16**	.15**	.01*	-.03**	.03**	-.05**	-.41**	-.55**	-.47**	1	
13. White	-.05**	.00	-.01	.01	-.04**	.09**	.05**	-.04**	.04**	-.10**	.16**	-.05**	1

*** p<.001; ** p<.01; *p<.05

Chapter 3

Developing and Validating the Montana Juvenile Probation Risk Screener

The criminal justice system has experienced a significant increase in the use of actuarial risk assessments during the past several decades (Miller & Maloney, 2013). Such assessments have been recognized as a key component of criminal justice reform and evidence-based decision-making for professionals in the adult and juvenile justice systems (Bushway, 2013; National Center for Juvenile Justice, 2006). With the pervasive use of risk assessments, methods for creating and validating them have become more sophisticated (e.g., Duwe & Freske, 2012). Such fine tuning increases accuracy and usability for these tools in the field. The primary research objective in the current investigation is to create and validate a new risk assessment instrument for juvenile probation officers in Montana based on current methods and localized data.

Montana juvenile probation currently uses the Back on Track (BOT) risk and needs assessment. The BOT was created in 1998 by the Washington State Institute for Public Policy and the Washington Association of Juvenile Court Administrators and Assessments. When the BOT was created, it was known as the Washington State Juvenile Court Assessment (WSJCA). In addition to being utilized in Montana, Washington State's instrument was adapted by Florida, which created the Positive Achievement Change Tool (PACT) and Vermont, which utilizes the Youth Assessment and Screening Instrument (YASI). All of the above-itemized instruments are comprised of a pre-screen risk assessment and a full needs assessment. The pre-screen assessment is a shortened version of the full assessment. It indicates whether a youth is at low, medium, or high risk to reoffend within the following year. Youth who receive a medium or high-risk score are given the full assessment. The full assessment is meant to identify risk and protective factors capable of guiding decisions about which services are most needed for rehabilitation.

A validation report of the BOT (McKay, Hollist, Bunch, Acton, Tillman, & Harris, 2015) found the tool was accurate but was not being used with fidelity across all Montana juvenile probation districts.

A primary barrier to using the tool with fidelity is the time it requires. The pre-screen BOT is comprised of approximately 40 risk factors. Information about risk factors is collected during interviews with the youths' parents or guardians, and school officials. Additionally, some information is gathered from the Juvenile Court and Assessment Tracking System (JCATS) if the youth has previously been cited for an offense. The time requirement to administer the pre-screen BOT is seen as a barrier to using the tool with fidelity. Additionally, the pre-screen BOT was adopted from Washington State verbatim. While it is common practice to adopt a validated instrument (e.g., Hamilton et al., 2016), predictive accuracy does not always translate from one state to the next. The current investigations attempts to address these two problems by creating a shorter risk assessment from localized data, seeking to maintain, or increase prediction accuracy with a tool that can be used with fidelity.

Risk Assessments in the Juvenile Justice System

Researchers have been studying formal prediction methodologies in the criminal justice system for almost a century. In 1928, Ernest W. Burgess created one of the first risk assessment instruments using what would later be called the Burgess Method (Burgess, 1928). This is a linear additive model that looks at several risk-predicting variables. For each risk variable that applies to an individual, one point is added to their total score. Thus, the more points an individual scores on the instrument, the more likely the individual is to act out the risk behavior being predicted (e.g., recidivate). Since the creation of the Burgess Method, researchers have been examining ways to increase accuracy of a risk assessment by finding both alternate models that predict risk and ways to add meaningful weight to risk-predicting variables. Andrews, Bonta, and Wormith (2006) describe four generations of risk assessment. The first generation is now commonly called "clinical judgment," or "clinical assessment." This generation relied on experience, knowledge, and intuition to assess the risk a youth may pose. The second generation was marked by creation of the actuarial risk assessment instrument (Burgess Method). The actuarial risk assessment facilitates consideration of a series of risk factors to determine the amount of risk, instead of relying on professional judgment alone. An actuarial instrument is preferable than clinical judgment, because it allows for more reliable, consistent, and unbiased judgment (Hilton, Harris & Rice, 2006; Bishop & Trout, 2002; Wilcox, Beech, Markall, & Blacker, 2009).

The first actuarial risk assessment instrument almost exclusively weighed static risk factors. Static risk factors are historical characteristics of the youth that cannot be changed, such as age at first offense and gender. First-generation risk assessment instruments were able to discern high and low risk, but, because it allowed for an evaluation of only static risk factors, it was unable to help facilitate an exploration of intervention strategies.

Third generation risk-assessment methodology includes both static and dynamic risk factors in the actuarial risk assessment instrument. Dynamic risk factors are factors that can potentially be changed with intervention. Examples of dynamic risk factors are friends, school performance, and activities the youth participates in (Van der Put, Dekovic, Stams, Van der Laan, Hoeve, & Amelsfort, 2011). Including dynamic risk factors in the actuarial risk assessment instrument opens the door to insights about potential intervention strategies that could lower a youth's risk.

Fourth generation risk-assessment instruments include protective factors. Protective factors are positive factors in a youth's life that are negatively correlated to recidivism. Examples of protective factors include a positive family or friend influences, structured activities, and academic success. All youth have risk and protective factors that push and pull them into delinquency. Both are important to recognize when determining the risk that a youth may pose. In addition to the inclusion of protective factors, the fourth-generation risk assessment is designed to link youth with a case management plan capable of addressing specific needs (Andrews, Bonta, & Wormith, 2006).

The evolutions of risk assessment coincides with the movement towards outcome-based systems concerned with recidivism reduction (IBM strategic assessment, 2004). One influential model developed to reduce recidivism includes the use of three guiding tenets (1) risk; (2) need; and (3) responsivity (Andrews, Bonta, & Hoge, 1990). This methodology is commonly referred to as the risk-need-responsivity (RNR) model (see Andrews & Bonta, 2010; Andrews, Bonta, & Hoge, 1990; Cohen, Christopher, Lowenkamp, & Robinson, 2018; Lowenkamp, Johnson, VanBenschoten, Robinson, & Holsinger, 2013). The *risk principle* is concerned with determining the youth's level of risk so appropriate services can be matched to the individual. The *need principle* focuses on dynamic criminogenic risk factors or those risk factors in the offender's life that can be changed with intervention. Finally, the *responsivity principle* constitutes the idea that treatment interventions and modalities are matched to the offender's learning styles and abilities. Currently in Montana juvenile probation, the RNR model is available through the use of the prescreen BOT, which measures risk and the full BOT, which measures need. However, as discussed earlier, districts are not using these tools with fidelity which reduces the RNR models effectiveness in Montana.

Methods

Sample

The dataset used for this study is queried from the Montana Juvenile Court Assessment Tracking System (JCATS). JCATS is Montana's juvenile justice data repository. All intakes from January 1, 2010 to December 31, 2015 that had a full BOT assessment are included in the dataset (*N* = 3,121).[11] Full BOT data are necessary to examine factors not found on the pre-screen BOT for possible inclusion on the Montana Juvenile Probation Screener (MJPS).

The eligible sample is then randomly divided into two groups, an Estimation sample and a Validation sample.[12] The Estimation sample is used for the creation of the MJPS. The Validation sample, meanwhile, is called upon to determine the MJPS's validity consistent with the literature (e.g., Tollenaar, & van der Heijden, 2013). Table 2.1 presents the demographic characteristics for the two samples.

---

[11] Certain youth were included in the sample more than once if they were cited with additional offenses within the study timeframe. The analyses in this investigation were run with and without duplicate cases to determine if these cases impacted the outcome. When duplicates were removed the prediction accuracy of the created screener decreased from when duplicates were included and is why this study utilizes all cases to create the risk screener.
[12] To randomize group selection, all youth were assigned a number through a random number generator. Five hundred youth were then randomly selected based on their generated number and placed into the Validation sample. The remaining 2,621 youth were placed into the Estimation sample.

Overall, the samples (Estimation and Validation) appear to be similar regarding demographic comparisons.

Table 2.1: Background Characteristics (*N* =3121)

Demographic		Estimation Sample	Validation Sample
Age	At Offense	14.99 (SD=1.63)	15.042 (SD=1.56)
Gender	Female	884 (33.7%)	199 (39.8%)
	Male	1737 (66.3%)	301 (60.2%)
Race/Ethnicity	White	2076 (79.2%)	396 (79.2%)
	American Indian	229 (11.4%)	56 (11.2%)
	Asian	9 (0.3%)	0 (0.0%)
	African American	52 (2%)	11 (2.2)
	Hispanic/Latino	84 (3.2%)	22 (4.4)
	Other	101 (3.9%)	15 (3.0%)
Intake Year	2010	426 (16.3%)	67 (13.4%)
	2011	112 (17%)	90 (18%)
	2012	366 (14%)	70 (14%)
	2013	347 (13.2%)	50 (10%)
	2014	520 (19.8%)	110 (22%)
	2015	517 (19.7%)	113 (22.6%)

Measures

Recidivism. The outcome of interest in this analysis is recidivism. Recidivism is operationalized as a new citation within the risk period. The risk period is a one-year span of time starting the day after the youth's initial intake date. Recidivism is dichotomously coded where 1 indicates the youth had a qualifying recidivating offense and 0 otherwise. Qualifying recidivating offenses include technical violations, criminal contempt, misdemeanors, and felony offenses. Table A (in Appendix B) displays the initial offense and recidivating offense frequencies for both the Estimation sample and the Validation sample. City ordinance offenses and status offenses are included in Table A but are not included in the analysis as recidivating offenses. When only qualifying recidivating offenses are kept in the analysis, the Estimation sample has a recidivism rate of 38.4% (1007), and the Validation sample has a recidivism rate of 39% (195).

Risk Factors. All risk and protective factors identified in this analysis come from the full BOT instrument. A total of 246 risk and protective factors from 11 domains on the full BOT were initially included as eligible factors for the new risk assessment instrument. The 11 domains that make up the full BOT can be seen in table B (in Appendix B). For all risk factors on the full BOT where the option of "select all that apply" is available, each of the option choices were built as dichotomous "Yes" or "No" responses. These nuanced questions were broken down into several individual risk factors.[13]

Item Selection Criteria

Risk assessment instruments are comprised of items or risk factors that predict an outcome of interest. These items are a combination of both static and dynamic risk factors. Historically, items have been selected for inclusion for risk assessments based on one or more of the following: professional experience (Wormith & Bonta, 2018), adoption of risk factors used in other tools (Baird, 2018), bivariate (Latessa, Lovins, & Lux, 2018), and multivariate techniques (Cohen, Lowenkamp, & Robinson, 2018). Factors selected by professionals in the field and factors adopted from previously validated instruments must be empirically examined to determine their ability to predict the outcome. "At a minimum, a significant bivariate association with recidivism is needed to identify an empirical relationship, or to make a determination that the item is a criminogenic predictor." (Hamilton, Kigerl, Campagna, Barnoski, Lee, Van Wormer, & Block, 2016 p. 233). Baird (2009) noted that items not meeting this minimum criterion will not add to the prediction of recidivism and could in fact reduce the assessments ability to predict outcomes. One of the more rigorous approaches for identifying risk factors utilizes multivariate techniques. An item may be a strong predictor at the bivariate level but may lose its impact on predicting

[13] For example, in Domain 7A, "Family history," youth are asked whether anyone living in their household had been jailed or imprisoned within the previous three months. Answer options to this question include: 1) Mother/female caretaker; 2) Father/male caretaker; 3) Older sibling; 4) Younger sibling; 5) Other family member; and 6) No jail/imprisonment history in family. Because this is a 'select all that apply' question, the six options were dichotomized in to separate variables, each with a "Yes" =1 or "No"=0 response (e.g., mother jail/imprisonment history, "Yes"/"No;" father jail/imprisonment history, "Yes"/"No," etc.).

the outcome when other factors are held constant. Such a finding indicates factors are indirectly predicting recidivism and are no longer necessary once other factors are held constant.

Several studies comparing strategies for selecting items for inclusion on risk assessment instruments for the prediction of a binary outcome (recidivism) have concluded that multivariate strategies utilizing logistic regression have outperformed instruments created utilizing other popular methods (Duwe, 2017; Hamilton, Kigerl, Campagna, Barnoski, Lee, Van Wormer, & Block, 2016; Hamilton, Neuilly, Lee, & Barnoski, 2014; Tollenaar & van der Heijden, 2012; Williams et al., 2014).

Validation Statistics

Risk assessment accuracy is an imperfect measurement and must be considered in a relative context. A risk assessment may be determined as "accurate" in the justice system, while not considered "accurate" in the medical sciences. At the same time, one measurement capable of determining accuracy does not exist. For this reason, it is important to run a variety of tests examining accuracy. The following strategy provides the ability to effectively determine prediction performance on a relative scale in the justice system.

The common statistic measuring risk assessment accuracy is derived from the Receiver Operating Characteristic (ROC) analysis, known as the area under the ROC curve (AUC; see Hanley & McNeil, 1982). The ROC curve is created by plotting the sensitivity (true positive rate) by 1 – specificity (false positive rate) across various cut-points used to classify youth as a recidivist or a non-recidivist. (Mossman, 2013). An AUC of 0 indicates perfect negative prediction, 0.5 indicates no better than chance prediction, and an AUC of 1 indicates perfect positive prediction (Van der Put, Van Vugt, Stams, & Van der Laan, 2013). AUC interpretations vary based on the field where they are used. It is common in the criminology and psychology risk assessment literature that an AUC score of 0.7 or above indicates strong prediction performance, between 0.6 and 0.7 indicates moderate performance, and anything below 0.6 indicates poor performance (Barnoski, 2004: Douglas, & Reeves, 2010; Mossman, 2013). Contemporary validation studies of popular risk assessment in the United States have found AUC's from 0.65 to 0.75 (see Singh, Kroner, Wormith, Desmarais, & Hamilton, 2018). ROC analysis also provides a test of statistical significance. This statistic determines if the AUC derived in ROC analysis is statistically different from chance performance in the population.

Analytic Strategy

The following analysis is broken into two sections: (1) risk screener creation; and (2) risk screener validation. The creation of the risk screener utilizes the Estimation sample and the following process: selecting the items to be included in the screener, weighting the risk factors for use on the screener, and developing cut points where final scores will fall into low, medium, and high risk of recidivism.

The current investigation will use a multi-step approach for item selection from the full-BOT for inclusion on the risk screener. The following five steps demonstrate this approach: (1) bivariate regression, (2) individual domain logistic regression, (3) full-model logistic regression, (4) professional opinion, and (5) final logistic regression. Each step in this approach removes risk factors not found to be predictive of recidivism. The final product will include a series of risk factors that are found to be predictive of recidivism at the bivariate and multivariate level.

Once items have been selected, weights are assigned to each risk factor. The Burgess Method of weighting risk factors is used for this screener. Finally, cut points of low, medium, and high risk are determined based on a simple examination of the risk scores and their associated probabilities of recidivism. To differentiate between low and medium risk, the recidivism base rate or average recidivism rate will be used as a barrier. Low risk will represent those youth that are lower than the average case to have a recidivating offense. Next, medium risk will encompass those scores that are between the base rate and a score that comes close to 50% of the youth having a recidivating offense. Finally, high risk will be the remaining scores in the distribution and represent youth that have a greater than 50% likelihood of having a recidivating offense.

The second section validates the newly created risk screener with both the Estimation and Validation sample. The Estimation sample is used to create the screener and it is assumed that because of this it will demonstrate inflated accuracy statistics. The Validation sample, however, is not used in the creation and is a more reliable source to conduct a retrospective validation of the screener.

Validation measures begin with basic descriptive cross-tabulations depicting the percent of recidivists and non-recidivist per risk score. At a minimum, there should be a basic pattern demonstrating an increase in recidivists as risk scores increase for both Estimation and Validation samples. Next, ROC analysis is conducted to formally compare the accuracy of the new risk screener to the prior pre-screen BOT. Finally, a logistic regression model will examine the risk level category's predictability of recidivism on the new screener and the pre-screen BOT while holding the effects of gender and race constant.

Results

Section 1: Risk Screener Creation

Item selection step 1: bivariate correlations. Utilizing the Estimation sample, bivariate relationships between individual risk factors and recidivism are investigated. Only risk factors found to have a statistically significant ($p \leq 0.05$) relationship to recidivism are kept for the next steps of the analysis. Correlation matrices based on domains are examined to assess the possibility of collinearity. It is expected that risk factors are highly correlated to the outcome variables (recidivism) but independent of each of the other risk factors. A risk factor that is highly correlated with another risk factor is an indication that both are measuring the same phenomenon and will not contribute uniquely to the overall risk score.[14] Additionally, risk factors were examined for directionality of correlation. Those risk factors

[14] Midi, Sarkar, and Rana (2010) suggest using 0.8 as an indicator of collinearity.

predicting in an illogical direction were removed from the analysis, following a strategy put forth by Hamilton et al., (2016) to maintain face validity. Upon examination, 94 risk factors were found to be too highly correlated ($r \geq 0.8$) to another risk factor, prompting them to be removed from the analysis. The determination of which risk factor was to be removed was based on an analysis of items with a lower bivariate correlation with the outcome variable.[15] Removing redundant risk factors through the process described here has been noted as a key to the creation of a successful risk assessment (Steinhart, 2009).

Table B (Appendix B) presents findings from the bivariate investigation. The first column, "Domain," specifies which domain each risk factor belongs to in the full BOT. The second column, "Bivariate Sig. Predictors," presents all risk factors demonstrating statistically significant correlations to recidivism. The third column, "r," presents the Pearson's r correlation coefficient. Eighty-one of the risk/protective factors on the full BOT are found to have statistically significant correlations at the bivariate level with recidivism. All non-statistically significant risk factors (61) are removed before proceeding with the analysis. Logistic regression is then used to further reduce the set of risk factors.

Item selection step 2: domain logistic regression. At this stage in the analysis, each domain on the full BOT has been reduced to only risk factors correlated at the zero order with recidivism and show low correlations with other risk factors. To further reduce the number of items, 11 individual logistic regression analyses were run, one for each domain, with recidivism as the dependent variable. Forty-eight risk factors maintained statistical significance in the domain logistic regression models, prompting them to be kept in the analysis. Table C (Appendix B) presents the factors that maintain statistical significance.

[15] An example of this comes from examining the variable "Not a victim of physical abuse," with a Pearson's r correlation coefficient of -0.881, alongside the variable "History of physical abuse by family member." The variable, "History of physical abuse by family member," was removed, because it has a lower correlation to recidivism than "Not a victim of physical abuse."

Item selection step 3: full logistic regression. Next, the 48 risk factors located from the domain logistic models are placed into a single logistic regression model.[16] After running the first full logistic regression with 48 factors, ten factors that reached a $p \leq 0.1$ were kept for the next model.[17] These ten factors are then included into an additional logistic regression model. Results of this logistic regression model show that six factors remained statistical significance ($p \leq 0.05$). Results from this regression are shown in Table D (Appendix B) in the column titled "First Model." The remaining six factors are then included in another logistic regression model. Results in Table D under the column titled "Second Model" show all six factors maintain statistical significance with recidivism. These final factors resulted in two protective factors (negatively correlated to recidivism and four risk factors (positively correlated to recidivism).

Item selection step 4: professional opinion. To ensure face validity, the six risk and protective factors were presented to 10 chief juvenile probation officers in Montana. The 10 respondents were asked to comment on the list of risk factors, explain how they currently collect the data, and to identify potential items not included that theoretical and/or practical experience indicate should be included. Overall responses to these risk and protective factors was positive. However, officers explained that terminology in some items can be ambiguous. For example, one officer explained that the protective factor of *History of positive adult non-family relationships not connected to school or employment* is very broad and can be difficult to determine exactly who qualifies as a positive adult non-family relationship. Similar sentiments were expressed about *Current antisocial friends* and *Youth demonstrates the ability to problem solve*.

A common theme emerged in the discussion of items that should be included in the risk assessment. These recommendations were focused on past misdemeanor and felony offenses of youth.

[16] Agresti (2007) suggests as a rule of thumb that there be at least 10 outcomes (recidivating offenses) for every predictor in a logistic regression model. The number of outcomes clearly reach this guideline with 1,007 recidivists, 1,164 non-recidivists, and only 48 factors.
[17] The conservative boundary of $p \leq 0.1$ was used as a cut-point to leave room for error as some risk factors were very close to the arbitrary 0.05 threshold.

There are two risk factors on the BOT that measure past offenses misdemeanor and felonies: *Total Misdemeanors* and *Total Felonies*.

According to Table B the risk factor *Total Misdemeanors* has a statistically significant ($p \leq 0.001$) bivariate correlation of 0.320 with recidivism. *Total Felonies* also show a statistically significant ($p \leq 0.001$) but much weaker correlation (0.081) with recidivism. As detailed in findings itemized in Table C, *Total Misdemeanors* maintained statistical significance in the domain logistic regression model. However, *Total Felonies* was removed from the analysis during the domain logistic regression, as it no longer presented statistically significant results holding constant the effects of other variables in the model. To further analyze these two risk factors, they are added (forced) into the resulting factors found from the "Second Model" to test their impact of predicting recidivism when holding all other significant factors constant.[18]

Logistic regression model results with *Total Misdemeanors* and *Total Felonies* are shown in Table D in the column titled "Third Model" All prior factors from the Second Model maintain statistical significant with recidivism. Additionally, *Total Misdemeanors* is statistically significant ($p \leq 0.001$) indicating this factor can be added to the risk screener. *Total Felonies*, however, was not statistically significant. The findings show that youth with past felonies are less likely to commit a recidivating offense holding all other factors constant. These results support the inclusion of *Total Misdemeanors* but fail to support the inclusion of the *Total Felonies*.

Item selection step 5: final logistic regression. A final logistic regression model was run without *Total Felonies* and is presented in Table 5 in the column titled "Final Model." All seven factors in the final logistic model present statistically significant results both at the bivariate and multivariate level and will construct the new risk screener.

---

[18] The strategy of forced inclusion was employed by Hamilton et al., (2016) in the creation of the STRONG-R risk assessment.

Finalizing items for risk screener. Two hundred and forty six risk factors were analyzed and reduced to seven in the new model to predict juvenile recidivism one year from initial intake. Protective factors that decreased the likelihood of committing a recidivating offense are reverse coded for the risk assessment. This simplifies the tool and increases usability. A draft of the final risk assessment tool is displayed as Figure A in Appendix B.

Weighting risk factors. The Burgess Method of weighting risk factors is used to add risk scores for each risk factor on the MJPS. The Burgess Method produces a simple cumulative risk score by adding one point to the total risk score for each risk factor that applies to a youth.

As discussed earlier, a variety of techniques have been employed to maximize risk assessment accuracy by obtaining meaningful weights for risk factors. One common technique is to use the factor's standardized regression coefficient obtained in logistic regression to inform risk factor weights (Gottfredson & Snyder, 2005; Williams et al., 2014; Silver, Smith, & Banks, 2000). Findings indicate, however, that advanced techniques do not significantly outperform simple strategies such as the Burgess Method (Grann & Langstrom, 2007; Gottfredson & Snyder, 2005; McKay, 2012; Silver, Smith, & Banks, 2000; Simon, 1972). While "meaningful weights" may optimally predict a criterion of interest, it was found that adding equal weight to risk factors, which creates an "improper linear model," are robust in making predictions and superior to clinical intuition (Dawes, 1978).

The Burgess Method was chosen for this screener because it is a proven strategy that is logical for the probation officers who will be using the screener. The transparency of scores maintains face validity and provides users with the easiest form of scoring which in turn increases inter-rater reliability. Advanced weighting techniques that yield obscure risk scores can easily be calculated and interpreted with computer automated scoring but could decrease face validity when completed by probation officers by hand (Hamilton et al., 2016). The level of Service Instrument (Wormith & Bonta, 2018), the Ohio Risk Assessment System (Latessa, Lovins, & Lux, 2018), and the Self-Appraisal Questionnaire (Loza,

2018) are three poplar and validated risk assessments that continue to utilize the Burgess Method for weighting risk factors.

Risk scores range from 0 (for youth without risk factors) to 7 (for youth that have all seven risk factors). There is a relatively even distribution of scores from 0 to 7 for both samples with the exception of 6 and 7, which garnered a lower number of youth. The distribution of risk scores for the Estimation sample and Validation sample are similar (see Table E in Appendix B). This is expected if findings from the Estimation sample can be generalized to a sample not used in the creation of the tool.

Creating cut points. Table F (Appendix B) is a cross-tabulation of each risk score and the percent and frequency of youth that had a recidivating offense. With each point increase in risk score, there is an almost 10 percentage point increase in the rate of recidivism for all risk score categories except for the risk score of 7 which has slightly lower but similar recidivism rate to the risk score of 6. This distribution is used to determine the cut-points of low, medium, and high risk. While cut points are created they are somewhat arbitrary. It is recommended that probation officers consider the actual risk score (0 to 7) when making their final decision.

To differentiate between low and medium risk, the recidivism average is used as a barrier. The average for the Estimation sample is approximately 38%. The average value most closely matches a risk score of 3. As shown in Table F, all scores below 3 are less likely than the average case to have a recidivating offense and will be designated in the low-risk category. When combined into one category, the low-risk category has only a 19.7% failure rate[19] with the majority (80.3%) showing success in the year-long period of risk. The medium-risk category will encompass the risk scores of 3 and 4. As discussed above, 3 is right at the recidivism base rate and the risk score of 4 shows that just over 50% of the youth who received a 4 had a recidivating offense. When combined into one category, youth indicated to have medium risk have a 44.9% failure rate and a 55.1% success rate. The remaining scores (5 through

[19] Failure rate in this context is referring to recidivism rate. Success, meanwhile, is referring to the absence of a recidivating offense in one year.

7) will be placed into the high-risk category. Young people receiving a high risk score show a failure rate of 65.9% and a success rate of 34.1%.

As discussed above, probation officers should consider the actual risk score (0 through 7) in their decision making and use these cut points as suggestions. A score of 0 should be considered as the *lowest-low* risk of recidivism with only 10.9% failure rate compared to a 2 which should be considered a *high-low* risk of recidivism which doubles the risk of those in the 1 category with a 28.4% failure rate.

Section 2: Risk Screener Validation

Recidivism rate by risk score examination. The Validation sample is called upon to act as a retrospective test on the MJPS because it was not used in the creation of the screener. Table G (Appendix B) displays the cross-tabulation of risk scores and whether or not the youth had a recidiviating offense for the Validation sample.

Similar to the results from the Estimation sample, the Validation sample shows an increasing failure rate as the risk score increases. Those youth who scored a low risk had a failure rate of 21.8% and a success rate of 78.2%. Those youth who scored a medium risk present a failure rate of 51.9% and a success rate of 48.1%. Finally, those youth who received a high risk score show a failure rate of 60.6% and a success rate of 39.4%. The rising failure rate as risk level increases is a positive sign that the risk score is a valid predictor of recidivism on a sample that was not included in the creation of the tool.

Receiver operating characteristic analysis. Table 2.2 displays the findings from ROC analysis. As discussed above, AUCs above 0.7 are indicative of an accurate and comparable risk assessment in the juvenile justice system. Results show an increase in prediction accuracy from the pre-screen BOT risk score to the MJPS score for bother Estimation and Validation samples. In the Estimation sample, the pre-screen BOT risk score results in an AUC of 0.685 and the MJPS score has a calculated AUC of 0.735 a 0.05 increase in AUC prediction accuracy. The difference between these two AUCs is statistically significant ($p \leq 0.001$).[20] The MJPS shows only a slight drop in AUC from the full Estimation sample to the full Validation sample (0.735 and 0.729 respectively) which is an additional indication that accuracy is maintained on an independent sample.

To further investigate the MJPS risk score, the Validation sample is broken down into four subsamples: white youth, American Indian youth, female, and male. As indicated in Table 2.2, the screener's prediction accuracy increases for the full sample and all subsamples over the accuracy of the pre-screen BOT. The MJPS risk score consistently shows AUCs in the 0.7 and up range, which is evidence of a strong predictor of recidivism. All AUCs were found to be statistically significant. While improvements are shown, no AUC for the MJPS score is statistically different from the pre-screen score in the Validation sample.

Table 2.2 ROC Analysis (AUCs)

Estimation Sample	Pre-Screen BOT Score AUC	MJPS Score AUC	AUC Improvement
Full (N=2621)	.685	.735	+.05
Validation Sample	Pre-Screen BOT Score AUC	MJPS Score AUC	AUC Improvement
Full (N=500)	.688	.729	+.041
White (n=396)	.658	.726	+.068
American Indian (n=56)	.723	.754	+.031
Female (n=199)	.666	.742	+.076
Male (n=301)	.698	.717	+.019

all AUCs significant at $p \leq .001$

Risk level category logistic regression. Next, logistic regression is used to evaluate risk level category (RLC) for the pre-screen BOT and MJPS. For this analysis, two logistic regression models are run on the Estimation and Validation sample. One model is constructed for the pre-screen BOT risk score and another for the MJPS score. Each model is predicting the outcome of a recidivating offense and consists of five binary independent variables: (1) Low risk; (2) Medium risk; (3) High risk; (4) Race (white) and (5) Sex (male). This procedure allows for an accuracy analysis of each RLC while holding

[20] Utilizing a difference in areas between two ROC curves calculation (see Hanley & McNeil, 1982).

constant the effects of race and sex. Table H (Appendix B) presents the results of the two logistic regression models for the Estimation sample.

Results show that youth with a medium risk score on the BOT pre-screen are 3.73 times more likely than low-risk youth to have a recidivating offense, and high-risk youth are 4.8 times more likely to have a recidivating offense. White youth are 23% less likely to have a recidivating offense than non-white youth and males are no more likely to have a recidivating offense than females.

Similar results are found for the MJPS risk scores. Having a medium risk on the MJPS, youth are 3.28 times more likely to have a recidivating offense compared to youth who have a low risk score, and youth with a high risk score are 7.67 times more likely to have a recidivating offense compared to youth with a low risk score. White youth and males are no more likely to have a recidivating offense compared to their respective counterparts.

Table I (Appendix B) displays the results from logistic regression models for the Validation sample. On the pre-screen BOT, youth with a medium risk score are 2.83 times more likely to have a recidivating offense compared to low-risk youth and youth with a high risk score are 5.26 times more likely to have a recidivating offense compared to low-risk youth. White youth are 43.1% less likely to have a recidivating offense compared to non-white youth and males are slightly more likely to have a recidivating offense.

The MJPS risk score shows that youth with medium risk scores are 3.66 times more likely to have a recidivating offense compared to low-risk youth, and youth with high risk scores are 5 times more likely to have a recidivating offense compared to youth with low risk scores. White youth are 46.6% less likely to have a recidivating offense compared to non-white youth and males are slightly more likely to have a recidivating offense compared to their female counterparts. On both the pre-screen BOT risk score and the MJPS risk score, logistic regression model findings show an increase in likelihood of recidivism for each point increase in risk score, while holding race and gender constant for both the Estimation and Validation samples.

Through each validation measure, the MJPS risk score demonstrates comparable, and at times, increased prediction accuracy over the pre-screen BOT. These findings provides evidence that the newly created MJPS is a valid predictor of risk and maintains, if not increases, prediction accuracy over the pre-screen BOT in a retrospective analysis for both Estimation and Validation samples.

Discussion

This investigation sought to accomplish three primary goals: (1) create a juvenile probation risk assessment from localized data; (2) reduce the number of risk factors from the pre-screen BOT; and (3) maintain/increase accuracy in predicting juvenile recidivism from past risk assessments. The final product should be able to replace the pre-screen BOT to locate risk as the first step of the risk-needs-responsivity (RNR) model.

The Montana Back on Track (BOT) was adopted verbatim from a tool developed in the State of Washington. Such an adoption process presents problems with accuracy and face validity. Montana and Washington are different states in many ways, and a generalized tool created in Washington may not maintain the same accuracy when used in a population that it was not created in. This investigation utilized six years of Montana juvenile justice system data with a total of 3,131 youth to both create and validate the new risk assessment instrument. Such a methodology satisfied the first goal of this inquiry.

The second primary goal of this investigation was to decrease the number of risk factors from the pre-screen BOT and, in turn, to decrease the time it takes for a probation officer to administer the tool. Accomplishing this goal could trigger an increase in administration fidelity, an issue discovered during the previous validation of the pre-screen BOT (McKay et al., 2015). Using bivariate regression, a series of logistic regression models, and professional advice from Montana probation officers, the new risk assessment titled the Montana Juvenile Probation Screener was created. The MJPS is comprised of only seven risk factors that each contribute uniquely to the prediction of recidivism. Each risk factor is correlated at the bivariate level and maintain statistical significant in a multivariate model. The cumulative findings indicate that the second goal set forward for this inquiry was satisfied.

The final goal of this investigation was for the newly created risk assessment (MJPS) to maintain and potentially increase accuracy and performance when compared to the pre-screen BOT during a one-year period. Accuracy was analyzed using a variety of methods: (1) risk score distributions; (2) risk level category logistic regression; and the most common analysis in the risk assessment field (3) receiver operator characteristic analysis (ROC). Through each analysis, the MJPS was found to be an accurate predictor of recidivism. Moreover, for each sample analyzed (Validation, Estimation, race, and gender) the MJPS outperformed the pre-screen BOT. These analyses provide sufficient evidence to confirm that the MJPS is an accurate predictor of recidivism, compared to the pre-screen BOT and demonstrates accuracy similar to other instruments used in the juvenile justice field.

The creation and validation of the MJPS is the product of methods identified in the current literature on juvenile risk assessment. However, as discussed by Bushway (2013) and Hamilton et al. (2016), methods used to create and validate popular risk assessments around the United States typically go unpublished. These methods are commonly claimed to be proprietary information or are located in agency reports that may not be available to the public or are otherwise difficult to locate. The current investigation not only sought to develop a risk assessment screener but to explicitly describe the steps in the creation of the tool to add to the growing knowledge in this field. The MJPS must serve as a baseline for future research and enable incremental improvements to be made. While this inquiry successfully accomplished the three primary goals articulated at its onset, findings presented here are not without cautions and limitations. Limitations are discussed below.

Limitations

Data collected for the analysis are based solely on youth who had a full BOT administered to them. Because less than 50% of youth with JCATS-recorded intakes were administered a full BOT, this data may not be representative of all youth in Montana's juvenile justice system. While this is a limitation, the large sample size used for this analysis provides a certain level of confidence to support the validity of the MJPS. Similarly, factors on the BOT are the only risk factors that this study could analyze. These factors are not exhaustive, and important risk factors for predicting recidivism may not have been included in the pool of potential factors.

Certain risk factors on the new assessment have been deemed by probation officers as difficult to define and, therefore, difficult to collect accurately. This could trigger challenges with inter-rater reliability. Inter-rater reliability is the ability for different practitioners to calculate the same risk score on the same youth. For example, one probation officer might decide that a youth does not have antisocial friends because the youth's friends are not in the JCATS system. Another probation officer, meanwhile, might decide the same youth does have antisocial friends because the youth's parents say they associate with antisocial friends. Training probation officers on the specific protocol of locating this information could help decrease inter-rater reliability, which, in turn, increases the accuracy and face validity of the instrument.

An additional limitation to this analysis is the inability to proactively test the validity of the new screener. All analyses were completed in a retrospective test with secondary data. There may be differences in accuracy levels when probation officers use this tool in the field. For this reason, it will be important to analyze MJPS validity in the future research.

The final limitation in this investigation is the unknown utility of the MJPS alongside the full BOT for juvenile probation in Montana. Implementation of the new tool will not work without probation buy-in. Several probation officers appear willing to use an updated screener. Others, however, have expressed hesitation. A focus on gaining buy-in from probation officers should be prioritized, if Montana chooses to adopt the MJPS.

Future Research

Future research should prioritize a prospective test of the MJPS. Internal validity, face validity, inter-rater reliability, fidelity of use, and usability were unable to be analyzed in this retrospective test of the screener. If a screener is not valid in all of these measurements, then it may not be beneficial for probation officers in the field.

The MJPS was intended to be a generalized tool, appropriate for use with males and females. There has been discussion of the possible need for assessments to be gender specific (Van Voorhis, Wright, Salisbury, & Bauman, 2010), however, with prior research demonstrating evidence that the two genders should be treated as separate populations (Else-Quest, Higgins, Allision, & Morton, 2012). Future investigations in Montana should determine if a gender-specific tool would benefit recidivism prediction among the juvenile-justice population. Hamilton et al., (2016) describe three methods for making assessments gender specific: (1) create different cut-points for males and females; (2) use gender as a risk factor where being male increases the final risk score and being female does not; and (3) separate males and females from the beginning of tool creation to locate different factors and weights for each sample. The MJPS demonstrates satisfactory accuracy for both males and females in the Validation sample but a deeper investigation is required to fully understand potential differences.

Similar to the discussion of gender-specific assessments, Kroner, Mills, and Reddon (2005) argue for offense-specific assessments. The MJPS specifically measures recidivism risk for technical violations, misdemeanors, and felony offenses. However, it may be more beneficial for an assessment to predict only violent offenses, drug crimes, or felony offenses. Future research should explore the possibility of creating a screener that can predict offense-specific outcomes through a more tailored approach.

Finally, future research should explore different methods of weighting risk factors and the possibility of automating the assessment. The Burgess Method of weighting risk factors was used for the assessment because it is the simplest weighting method, one that can be easily used and calculated by probation officers in the field. If the MJPS were automated, this would open the possibility of using more complex factor weights, which could therefore improve prediction accuracy.

Conclusion

The Montana Back on Track risk and needs assessment instrument was created to assist in service placement for youth on probation. Prior to the current investigation, it was discovered that the pre-screen BOT is a valid predictor of risk for first time offending youth in Montana. However, due to the lack of buy-in from many districts, most youth receive neither a pre-screen BOT nor a full BOT, and many districts do not see the utility of the instruments. The overall purpose of this investigation was to simplify the screening process by creating a shorter instrument that would allow probation officers to determine the risk level a youth poses in a shorter amount of time. A shorter instrument that maintained risk prediction accuracy could increase buy-in from probation officers and, in turn, increase fidelity and consistency. This inquiry's goal was largely accomplished. Out of the 246 risk factors on the full BOT, Montana data was used to isolate seven factors found to be important predictors of recidivism among justice-involved youth. With these seven risk factors, the brief screener was found to be valid and improved prediction accuracy over the pre-screen BOT in a retrospective test. A pilot test of the MJPS is recommended to further analyze the assessment's abilities in the field. The current investigation should be a starting point from which improvements can be made, rather than as a final product.

Appendix B

Table A: Initial and Recidiviating Offense Frequencies

Initial Offense	**Estimation Sample**		**Validation Sample**	
	Freq.	**%**	**Freq.**	**%**
City Ordinance	7	0.3%	1	0.2%
Status	441	16.8%	74	14.8%
Other Non-Misdemeanor	2	0.1%	2	0.4%
Technical/Criminal Contempt	105	4.0%	16	3.2%
Misdemeanor	1789	68.3%	370	74.0%
Felony	276	10.5%	37	7.4%
Most Serious Recidivism (1 Year)	1146	43.7%	221	44.2%
City Ordinance	1	0.1%	1	0.5%
Status	137	12.0%	25	11.3%
Other Non-Misdemeanor	1	0.1%	0	0.0%
Technical/Criminal Contempt	84	7.3%	12	5.4%
Misdemeanor	797	69.5%	157	71.0%
Felony	126	11.0%	26	11.8%

Table B: Bivariate Regression: Pedictors of Recidivism

Domain	Bivariate Sig Predictors	r
Record Referrals	TotalMisd	.320**
	AgeFirst	-.223**
	ConfinedDetention	.203**
	MisdPerson	.188**
	TotalFel	.081**
	TotalWeap	.076**
	FailureToAppear	0.058
School	Performance	-.263**
	LikelihoodStaying	-.257**
	Conduct	-.235**
	YouthInvolv	-.227**
	Attendance	-.221**
	YouthCloseWith	-.218**
	ExpulsionSuspension	.180**
	NoSpecialEd	-.072**
	Alternative	.069**
	LearningDis	.064**
	CogDel	-.057**
	GED	.044*
	College	-.044*
Free Time	CurrentUnstrucActivity	-.219**
	CurrentStucActivity	-.214**
	HobbyGroup	-.112**
	Volunteer	-.094**
	CommCulture	-.072**
Employment	HistorySuccessEmploy	-.231**
	HistoryPosWorkRel	-.257**
	UnderstandsJob	-.220**
	CurrentPosWorkRel	-.179**
Relationships	AdmiresAntiPeers	.263**
	CurAntiSocFriends	.217**
	CurrentCommunityTies	-.188**
	HistNonFamRelationship	-.168**
	RomaticRelationship	.077**
	ProSocFriends	-.125**
Current Living Situation	ParentalAuthority	.268**
	RunawayKickedOut	.242**
	ApprPunishment	.204**
	FamAcitivity	-.202**
	FamilyConflict	.188**
	Income	-.180**
	OutHomePlacement	.178**
	ParentalSupervision	.177**
	NoOneJail3Months	-.176**
	SupportNetwork	-.170**
	NoProbParents	-.166**
	FamilySupport	-.170**
	CloseDad	-.111**
	LiveWithFather	-.101**

Domain	Bivariate Sig Predictors	r
Drug/Alc	NoHistDrugProbs	-.185**
	Treatment	.182**
	CurrentDrugEduProb	.161**
	NoHistAlcoholProb	-.156**
	CurrentDrugFriendProb	.151**
	CurrentAlcEduProb	.146**
	CurrentDrugUse	.144**
	DrugAlcAssessment	.142**
	CurrentDrugFamilyProb	.137**
	CurrentDrugCrimeProb	.134**
Mental Health	MentalHealthStatus	.161**
	HistoryMentalHealth	.122**
	HealthInsurance	-.110**
	VicNeglect	.107**
	NoPysicalAbuse	-.088**
	AbuseByOther	.048*
Attitudes Behaviors	Attitude	-.306**
	RespectProperty	.266**
	Optimism	.264**
	BeliefInConditions	.250**
	Impulsive	.248**
	Control	.231**
	Emotions	.203**
Aggression	BeliefFighting	.279**
	BeliefYelling	.273**
	Frustration	.257**
	ViewOfIntentions	.222**
	NoReportsViolence	-.164**
Skills	GoalSetting	-.301**
	ProblemSolving	-.281**
	DealingWithFeelings	-.257**
	DealingWithOthers	-.244**
	InternalTriggers	-.223**
	Thinking	-.208**

Table C: Logistic Regression: Predictors of Recidivism

Domain	Sig Logistic Predictors	Exp(B)
Record Referrals	TotalMisd	1.55
	AgeFirst	0.79
	ConfinedDetention	1.25
	-	-
	-	-
	TotalWeap	1.78
	-	-
School	Performance	0.81
	LikelihoodStaying	0.86
	Conduct	0.84
	YouthInvolv	0.86
	-	-
	YouthCloseWith	0.87
	ExpulsionSuspension	1.10
	-	-
	-	-
	-	-
	CogDel	0.11
	-	-
	-	-
Free Time	CurrentUnstrucActivity	0.77
	CurrentStucActivity	0.80
	HobbyGroup	0.71
	Volunteer	0.54
	-	-
Employment	-	-
	HistoryPosWorkRel	0.67
	-	-
	CurrentPosWorkRel	0.03
Relationships	AdmiresAntiPeers	1.53
	CurAntiSocFriends	1.90
	CurrentCommunityTies	0.81
	HistNonFamRelationship	0.87
	-	-
	ProSocFriends	0.70
Current Living Situation	ParentalAuthority	1.53
	RunawayKickedOut	1.20
	ApprPunishment	1.11
	FamAcitivity	0.78
	-	-
	Income	0.81
	-	-
	-	-
	NoOneJail3Months	0.72
	-	-
	-	-
	FamilySupport	0.79
	-	-
	-	-

Domain	Sig Logistic Sig Predictors	Exp(B)
Drug/Alc	NoHistDrugProbs	0.52
	Treatment	1.35
	-	-
	-	-
	CurrentDrugFriendProb	1.36
	CurrentAlcEduProb	2.02
	-	-
	-	-
	-	-
	-	-
Mental Health	MentalHealthStatus	1.87
	-	-
	HealthInsurance	0.69
	VicNeglect	1.42
	-	-
	-	-
Attitudes Behaviors	Attitude	0.70
	RespectProperty	1.10
	Optimism	1.23
	-	-
	Impulsive	1.16
	-	-
	Emotions	1.16
Aggression	BeliefFighting	1.40
	BeliefYelling	1.28
	Frustration	1.44
	-	-
	-	-
Skills	GoalSetting	0.74
	ProblemSolving	0.86
	DealingWithFeelings	0.82
	-	-
	-	-
	Thinking	0.77

Table D: Logistic Models

First Model	Exp(B)	Second Model	Exp(B)	Third Model (forced)	Exp(B)	Final Model	Exp(B)
AgeFirst	0.79***	AgeFirst	0.79***	AgeFirst	.82***	AgeFirst	0.83***
CurAntiSocFriends	1.70***	CurAntiSocFriends	1.71***	CurAntiSocFriends	1.56***	CurAntiSocFriends	1.34***
ProSocFriends	0.88	HistNonFamRelationship	0.88**	HistNonFamRelationship	.88**	HistNonFamRelationship	0.88**
HistNonFamRelationship	0.89**	RunawayKickedOut	1.20***	RunawayKickedOut	1.14***	RunawayKickedOut	1.14***
RunawayKickedOut	1.18***	BeliefFighting	1.36***	BeliefFighting	1.28***	BeliefFighting	1.3***
VicNeglect	0.95	ProblemSolving	0.76***	ProblemSolving	.79***	ProblemSolving	0.79***
Impulsive	1.05	-	-	TotalMisd	1.35***	TotalMisd	1.34***
BeliefFighting	1.32***	-	-	TotalFel	0.90	-	-
Optimism	1.12	-	-			-	-
ProblemSolving	0.81***	-	-			-	-
-	-	-	-			-	-

***p ≤ .001; ** p ≤ .01; * p ≤ .05

Table E: Distribution of Risk Scores for Estimation and Validation Samples

Estimation Sample			Validation Sample		
Risk Score	*n*	%	Risk Score	*n*	%
0	331	12.63	0	63	12.60
1	436	16.63	1	80	16.00
2	402	15.34	2	100	20.00
3	417	15.91	3	74	14.80
4	439	16.75	4	84	16.80
5	346	13.20	5	60	12.00
6	211	8.05	6	32	6.40
7	39	1.49	7	7	1.40
Total	2621	100.00	Total	500	100.00

Table F: Distribution of Risk Scores and Recidivism (Estimation Sample)

Risk Score	Recidivism (1 Year) No		Recidivism (1 Year) Yes		Total	Risk Level	Failure Rate	Success Rate
0	295	89.1%	36	10.9%	331			
1	356	81.7%	80	18.3%	436	**Low**	19.7%	80.3%
2	288	71.6%	114	28.4%	402			
3	258	61.9%	159	38.1%	417	**Medium**	44.9%	55.1%
4	214	48.7%	225	51.3%	439			
5	127	36.7%	219	63.3%	346			
6	63	29.9%	148	70.1%	211	**High**	65.9%	34.1%
7	13	33.3%	26	66.7%	39			

Table G: Distribution of Risk Scores and Recidivism (Validation sample)

Risk Score	Recidivism (1 Year)				Total	Risk Level	Failure Rate	Success Rate
	No		Yes					
0	58	92.1%	5	7.9%	63	Low	21.8%	78.2%
1	64	80.0%	16	20.0%	80			
2	68	68.0%	32	32.0%	100			
3	46	62.2%	28	37.8%	74	Medium	51.9%	48.1%
4	30	35.7%	54	64.3%	84			
5	26	43.3%	34	56.7%	60	High	60.6%	39.4%
6	10	31.3%	22	68.8%	32			
7	3	42.9%	4	57.1%	7			

Table H: Risk Level Category Logistic Regression (Estimation Sample)

Risk Factors	Pre-Screen (Exp(B))	MJPS (Exp(B))
Medium Risk	3.73***	3.28***
High Risk	4.8***	7.67***
White	0.77*	0.834
Male	1.12	1.150

***p ≤ .001; ** p ≤ .01; * p ≤ .05

Table I: Risk Level Category Logistic Regression Validation Sample

Risk Factors	Pre-Screen (Exp(B))	MJPS (Exp(B))
Medium Risk	2.83***	3.66***
High Risk	5.26***	5.00***
White	0.569*	0.534**
Male	1.36	1.47

***p ≤ .001; ** p ≤ .01; * p ≤ .05

Montana Juvenile Probation Screener

Rater ID #:__ __ __ __ __ __ Youth JCATS #: __ __ __ __ __ __ __ __ Intake Date:____/____/_______
MM / DD / YYYY

1. Age at First Offense

This is youth's first offense..0
13 or older..0 +
Under 13..1

2. Misdemeanor Referrals

One or zero total prior referrals.. 0 +
More than 1 referral.. 1

3. Positive Adult Role Model

Yes.. 0 +
No.. 1

4. Current Friends

No anti-social friends.. 0 +
Current anti-social friends.. 1

5. History of Running Away / Kicked out

No.. 0 +
Yes.. 1

6. Belief in Fighting

No.. 0 +
Yes.. 1

7. Problem Solving

Yes.. 0 +
No.. 1

Total Score.. =

Screener Indicated Risk:	0-2 Low Risk	3-4 Medium Risk	5+ High Risk
Probability of Recidivism:	*20%*	*45%*	*66%*

Practitioner's Judgement (*check*): ☐ Low Risk ☐ Medium Risk ☐ High Risk

If practitioner's judgement does not match indicated risk please briefly explain why risk should be higher/lower:____________________________________

____

____

____

Disposition:

☐ Formal Probation ☐ Informal Probation ☐ Diversion ☐ Dismissed
☐ Pending

Figure A: The Montana Juvenile Probation Screener (Example Format)

Chapter 4

Utilizing Conjunctive Analysis of Case Configuration to Weight Burgess Risk Scores

When developing a risk assessment instrument to predict the likelihood of a future outcome, researchers are confronted with several decisions that will impact the prediction accuracy of the final product. Defining the outcome variable that will be predicted is one of the first decisions that must be made. Take, for example, the outcome of recidivism. Researchers must define the period of risk (e.g., 30 days, 60 day, 365 days, etc.), which crimes constitute a recidivating offense (e.g., status, misdemeanor, or felony), and how their recidivating offense will become known (e.g., arrested, charged, found guilty, etc.). Next, risk factors which will comprise the final risk assessment must be located to predict the outcome. Researchers might select risk factors that are already being used on a different assessment or those detailed in the academic literature, create a panel of professional stakeholders who will identify risk factors based on their professional knowledge, or, more commonly, use data and statistical analyses to help determine which risk factors are the most predictive of the outcome in question. Once risk factors have been selected for the assessment, weight must be assigned to these factors. Weighting risk factors appears to be one of the least discussed processes in the risk assessment literature. This could be due to the fact that a myriad of weighting strategies have been investigated, but a single most accurate method for this process has not been located (see Gottfredson & Snyder, 2005; Grann & Langstrom, 2008). The following analysis adds to the current body of knowledge about risk assessment weighting by investigating a new strategy based on Conjunctive Analysis of Case Configuration (CACC).

Weighting Risk Factors

The debate about weighting risk factors as detailed in the academic literature on risk assessments has centered on the best strategy to maximize prediction accuracy and whether it is even necessary relative to simple methods. Burgess (1928) created the first justice system risk assessment, which was designed to predict success or failure for parolees. Success, in this context, is referring to an individual not having a new citation in a certain period of risk. Failure, alternatively, is referring to an individual having a new offense citation in this same period of risk. In the Burgess assessment, a series of risk factors are given a weight of 1 when present and 0 when absent. Once the assessment has been completed, each risk factor score is summed to a total score. The higher the total score, the higher the probability of failure while on parole. This linear additive model that assigns a simple 0 or 1 score to each risk factor, has become known as the Burgess Method. The Burgess Method is still widely used and popular due to in part to its transparent and simplistic scoring design, one that has been validated as accurate across several different studies (e.g., VanNostrad & Lowenkamp, 2013; Rice, Harris, & Lang, 2013; Barnoski & Drake, 2007; Andrews, Bonta, & Wormith, 2004).

The simplicity and transparency of the Burgess Method comes with inherent flaws. Duwe and Kim (2016) outline two primary limitations of the Burgess Method. First, the Burgess Method assumes that all risk factors equally predict outcome. The method gives an equal weight of 1 to all risk factors, regardless of whether certain factors are stronger predictors than others. Second, the method does not consider interrelationships among predictor items. If two items on the assessment are highly correlated, risk scores will be excessively inflated because having one risk factor increases the likelihood of having the other risk factor. An additional limitation not discussed in current literature is the implicit assumption that the combination of risk factors that comprise a final Burgess Risk Score is inconsequential to the likelihood of the outcome being predicted. For instance, in an assessment with seven risk factors weighted with the Burgess Method, a score of 3 can be obtained through several different combinations of risk factors. However, the tools assume that all assessments scoring a 3 will have an equal likelihood of the same outcome, regardless of these combinations. With the inherent flaws of the Burgess Method, researchers have attempted to add more meaningful weights to address these concerns (e.g., Gottfredson & Snyder, 2005).

Bivariate regression and logistic regression are two common statistical techniques used in an attempt to add meaningful weight to risk factors (Duew & Freske, 2012; Howard & Dixon, 2012; VanNostrand & Rose, 2009; Grann & Langstrom, 2007; Silver, Smith, & Banks, 2000; Gottfredson &

Snyder, 2005). Bivariate regression simply takes the regression coefficient calculated for each factor on the outcome variable (most commonly recidivism) using Pearson's r or bivariate logistic regression as the weight for the risk factor (Grann & Langstrom, 2007). While this technique adds "meaningful" weight at the bivariate level, it still does not consider the interrelationships of the predictor variables. Alternatively, a weighting strategy based on multivariate logistic regression attempts to address this limitation. This method begins by including all predictor variables as covariates in a logistic regression model with the outcome being predicted (e.g., recidivism) as the dependent variable. Using the derived logistic regression formula, predicted probabilities of the outcome for all cases in the sample are calculated and serve as the final risk score on the assessment.

Additional statistical techniques utilized to try and locate meaningful weights for risk factors include: multiple linear regression (Gottfredson & Snyder, 2005; Gottfredson & Gottfredson, 1980), iterative classification (Silver, Smith, & Banks, 2000; Steadman, Silver, Monahan, Appelbaum, Robbins, Mulvey, Grisso, Roth, & Banks, 2000), recursive partitioning (Silver, Smith & Banks, 2000), neural network model (Grann & Langstrom, 2007; Caulkins, Cohen, Gorr, & Wei, 1996), Multiple Models Tool (Silver & Chow-Martin, 2002), the Nuffield Method (Nuffield, 1982), clustering methods (Gottfredson & Snyder, 2005; Gottfredson & Gottfredson, 1980), multivariate contingency (Gottfredson & Gottfredson, 1980) discriminant analysis (Gottfredson & Snyder, 2005), bootstrap methods (Gottfredson & Snyder, 2005), and Machine Learning algorithms (Barnes & Hyatt, 2012; Berk, Sherman, Barnes, Kurtz, & Ahlman, 2009).

While researchers have employed a diverse selection of techniques to add meaningful weights to scores assigned to risk assessment items, no one method has stood out above the rest. The most advanced techniques for weighting risk factors still do not significantly outperform simple tests such as the Burgess Method (Gottfredson & Snyder, 2005), and some have been found to decrease prediction accuracy (Grann & Langstrom, 2007).

Conjunctive Analysis of Case Configuration

Common quantitative analytic techniques utilized to understand causality are often described as variable oriented (Ragin, 2013). The purpose of these techniques is to examine a set of independent variables across a sample collected from a population of individuals to identify the independent variables that explain the variation in an outcome or dependent variable. Linear regression or logistic regression are examples of such techniques. The effect of any independent variable in the model is assumed to be constant across different contexts. However, the effect that combinations of independent variables have on the dependent variable is often misrepresented in these techniques (Miethe, Hart, & Regoeczi, 2008). Case-oriented approaches attempt to examine these complex interrelationships and how the combination of independent variables impact the dependent variable.

Conjunctive analysis of case configuration (CACC) is one innovative case-oriented approach for examining how a combination of independent variables affect a dependent variable (Miethe, Hart, & Regoeczi, 2008). CACC begins by selecting the independent and dependent variables of interest and locating every possible combination of the independent variables that can be arranged. Each combination represents one case configuration. As a demonstration, if there are three dichotomously coded (0, 1) independent variables X1, X2, and X3, the following conjunctive matrix (Table 3.1) exhausts all possible combinations of independent variables.

Table 3.1: CACC Example 1

Config #	**X1**	**X2**	**X3**
1	0	0	0
2	0	0	1
3	0	1	1
4	1	1	1
5	1	0	0
6	1	1	0
7	1	0	1
8	0	1	0

Each row represents a "case configuration." Case configuration 1 incorporates those observations in the sample that are coded 0 for X1, X2, and X3. Case configuration 2 incorporates those observations in the sample that are coded 0 for X1 and X2 and coded 1 for X3. This continues down through all eight possible combinations.

The number of possible case configurations depends on the number of independent variables. For the table above there were three dichotomously coded independent variables, thus the total number of case configurations can be located with the following formula:

$$\text{Configurations} = 2^{\#\ of\ dichotomously\ coded\ variables)} = \ 2^3\ = 8$$

Five dichotomously coded independent variables would result in 32 configurations ($2^5\ = 32$). While less common, independent variables can have more than two categories. If there are both dichotomously coded variables and trichotomously coded variables, the following formula is used to calculate the total number of cases:

$$\text{Configurations} = 2^{\#\ of\ dichotomously\ coded\ variables} \times 3^{\#\ of\ trichotomously\ coded\ variables}$$

Once all cases have been located, CACC proceeds by aggregating each observation into their respective case configuration. The column titled "*n* Cases" in Table 3.2 presents the number of observations located for that unique case configuration. The final column "Y" is the proportion of cases that experienced the outcome of interest in each respective case configuration.

Table 3.2: CACC Example 2

Config #	**X1**	**X2**	**X3**	***n* _Cases**	**Y**
1	0	0	0	nc1	y1/nc1
2	0	0	1	nc2	y1/nc2
3	0	1	1	nc3	y1/nc3
4	1	1	1	nc4	y1/nc4
5	1	0	0	nc5	y1/nc5
6	1	1	0	nc6	y1/nc6
7	1	0	1	nc7	y1/nc7
8	0	1	0	nc8	y1/nc8

To further illustrate this process, Table 3.3 provides an example of actual variable names with fictional data. For this example, gender, race, and past arrests are dichotomously coded. The dependent variable is recidivism. Each case contains a different number of observations ranging from 10 observations to 97. Miethe, Hart, & Regoeczi, (2008) recommend removing any case configurations that have fewer than 10 observations. Fewer than 10 observations may lead to inaccurate or misleading proportions in the dependent variable (Miethe, Hart, & Regoeczi, 2008). Cases with 10 or more observations are referred to as "dominant" case configurations. Table 3.3 explores which combinations of independent variables increase or decrease the likelihood of recidivism.

Table 3.3: CACC Example 3

Config #	**Gender**	**Race**	**Past Arrest**	***n* _Cases**	**Recidivism**
1	Female	Non-White	No Past Arrest	81	0.17
2	Female	Non-White	Past Arrest	35	0.55
3	Female	White	Past Arrest	10	0.65
4	Male	White	Past Arrest	65	0.89
5	Male	Non-White	No Past Arrest	78	0.11
6	Male	White	No Past Arrest	68	0.21
7	Male	Non-White	Past Arrest	44	0.89
8	Female	White	No Past Arrest	97	0.17

Logistic regression is an appropriate and commonly used technique to analyze the effect of gender, race, and past arrests on the dichotomously coded recidivism. However, logistic regression and other similar "main effects" models assume contextual invariance for the independent variables' effects (Drawve, Thomas, & Hart, 2017). If logistic regression was run on the variables from Table 3.3 and the odds ratio of gender was found to be a 1.5, this could be interpreted that males are 50% more likely to commit a recidivating offense than females —provided that males are the reference category— across all combinations of remaining variables in the model (race and past arrests). Main effects models like logistic regression do not estimate the conjunctive or context-specific influences.

Contrary to logistic regression's assumption of contextual invariance, CACC assumes that the combination of independent variables (case configurations) influences the dependent variable. As shown in Table 3.3, the proportion of recidivism differs among case configurations. Certain configurations have higher proportions of recidivists (high of 0.89) and certain combinations have lower proportions of recidivists (low of 0.11). An examination could be made of each combination to determine what factors in combinations are causing high and low rates of recidivism. One structured approach to investigate risk factor combinations is through "paired comparisons" (see Drawve, Thomas, & Hart, 2017; Hart, Rennison, & Miethe, 2017).

Two studies provide an excellent example of the limitations of main effect models and the utility of conjunctive analysis (Drawve, Thomas, & Hart, 2017; Drawve, Thomas, & Walker, 2014). Utilizing logistic regression, Drawve et al., (2014) found support for major variables derived from Routine Activity Theory (Cohen & Felson, 1979) in predicting the likelihood of an aggravated assault arrest. However, using the same variables from the first investigation, Drawve et al., (2017) ran CACC in a follow-up study and found: "the nature and magnitude of contextual variability in this study suggests that any of these particular variables may enhance, inhibit, or have no discernable effect on the likelihood of arrest, depending on the particular situational context that is the focus of one's attention" (p. 130). While main effect models such as logistic regression can be useful in providing parsimonious representations of average variable effects across different contexts, findings may be oversimplified. CACC can be a useful tool to empirically evaluate the findings from main effect models by examining combinations of variables and their impact on the outcome being predicted.

Current Research

The purpose of this study is to develop a method for weighting risk factors on a risk assessment to significantly improve prediction accuracy over the simple Burgess Method (Burgess, 1928). Previous studies that have attempted to add meaningful weight to risk factors have largely failed to increase accuracy and, at times, have even observed a decrease in accuracy with more complex methods (see Grann & Langstrom, 2007). Research presented here investigates a method that has not been attempted, using conjunctive analysis of case configuration (CACC).

This research began by acknowledging the Burgess Method produces an accurate risk score but that there is room for improvement. Thus, instead of starting from nothing, this inquiry attempted to build on the Burgess Method by focusing on one of its main limitations: the Burgess Method does not account for the combination of risk factors that go into a risk score. Instead, there are several ways that different combinations of risk factors can create the same risk score and these different combinations likely influence the outcome being predicted. CACC addresses this limitation.

CACC locates every possible risk factor combination and displays how these combinations affect the outcome. Probabilities of recidivism based on these different combinations are merged into the Burgess Method risk scores to create the new risk score, which is referred to as the "CACC Risk Score." Risk scores based on the original Burgess Method and risk scores derived from logistic regression are used as a source of comparisons against the CACC risk score. Risk scores derived from logistic regression are selected as a comparison scores due to the frequency in which this method is found in the literature (see Grann & Langstrom, 2007; Silver, Smith, & Banks, 2000; Gottfredson & Snyder, 2005). The current investigation hypothesizes that risk prediction accuracy in the Burgess Method can be significantly improved by accounting for the effect that different combinations of risk factors have on the predicted outcome.

Methods

Sample

Data for this study came from the Montana Juvenile Court Assessment Tracking System (JCATS). JCATS is Montana's juvenile justice data repository. All intakes from January 1, 2010 to December 31, 2015 are included in the initial sample (N = 3,121). The collected sample is then randomly divided into two groups, an Estimation sample (n = 2,621) and a Validation sample (n = 500).[21] The Estimation sample is used to locate the weights on the risk assessment. The Validation sample,

[21] To randomize group selection, all youth are selected based on an assigned number through a random number generator.

meanwhile, is called upon to determine and compare the accuracy of the risk assessment and assigned weights.

The Montana Juvenile Probation Screener

The Montana Juvenile Probation Screener (MJPS) is used for the following analysis. The previous chapter detailed the process of developing the screener and validating for youth on probation in Montana. The MJPS is comprised of seven risk factors that were found to be highly predictive of recidivism in both bivariate and multivariate models. Factors that comprise the MJPS are: (1) Youth's first offense was under the age of 13, (2) Youth currently has antisocial friends, (3) Youth does not have a history of non-family adult relationships, (4) Youth has a history of running away or has been kicked out of the house, (5) Youth believes in fighting, (6) Youth does not problem solve, and (7) Youth has more than one misdemeanor offense. One point is added to the total score for each risk factor that applies to the youth. There are eight possible risk scores ranging from 0 to 7, where lower scores are indicative of lower recidivism risk and higher scores of higher recidivism risk. The original risk scores from the MJPS are used as a reference risk score, which is referred to as "Burgess Risk Score" from this point on, and utilized to determine if new weighting methods increase prediction accuracy over the original method.

Measures

Table 3.4 presents the demographics for the sample used in this study. The dependent variable is recidivism. Recidivism is defined as a new technical, misdemeanor, or felony citation within a one-year period of risk and is dichotomously coded (0, 1). The independent variables are all risk factors from the MJPS along with the demographic variables of race, age, gender and intake year. Each risk factor is dichotomously coded where 1 = the presence of the risk factor. Age and intake year are continuous variables. Gender is coded such that male = 1. Race is dichotomously coded where 1= white, 0 = non-white.

Table 3.4: Background Characteristics (*N*=3121)

Demographic		Estimation Sample	Validation Sample
Age	At Offense	14.99 (SD=1.63)	15.042 (SD=1.56)
Gender	Female	884 (33.7%)	199 (39.8%)
	Male	1737 (66.3%)	301 (60.2%)
Race/Ethnicity	White	2076 (79.2%)	396 (79.2%)
	American Indian	229 (11.4%)	56 (11.2%)
	Asian	9 (0.3%)	0 (0.0%)
	African American	52 (2%)	11 (2.2)
	Hispanic/Latino	84 (3.2%)	22 (4.4)
	Other	101 (3.9%)	15 (3.0%)
Intake Year	2010	426 (16.3%)	67 (13.4%)
	2011	112 (17%)	90 (18%)
	2012	366 (14%)	70 (14%)
	2013	347 (13.2%)	50 (10%)
	2014	520 (19.8%)	110 (22%)
	2015	517 (19.7%)	113 (22.6%)

Analytic Strategy

The analysis began by conducting conjunctive analysis of case configuration (CACC) on the Estimation sample using each risk factor as independent variables and recidivism as the dependent variable. This analysis serves two purposes. First, results from CACC are used to illustrate one of the limitations of the Burgess Method discussed above. CACC locates how many combinations (case configurations) make up each risk score (0 through 7) on the MJPS and then displays the variation in recidivism based on each risk score. Second, probabilities of recidivism for each case configuration are used in conjunction with the original Burgess Risk Score to construct new risk scores (CACC Risk Scores).

Risk scores for each case in the sample are calculated by multiplying the original Burgess Risk Score by the probability of recidivism found for the unique case configuration the youth belongs to from CACC. The formula for calculating the CACC Risk Score is displayed below:

$$CACC\ Risk\ Score = Burgess\ Risk\ Score\ \times\ CACC\ Probability\ of\ Recidivism$$

To demonstrate, youth A and youth B both scored a 4 on the MJPS. Youth A belongs to a case configuration that has a probability of recidivism at 0.23 and Youth B belongs to a case configuration that has a probability of recidivism at 0.45. Youth A’s new score is 0.92 (4 × .23) and Youth B’s new score is

1.8 (4 × .45). The purpose of this strategy is to include a measurement that adjusts for the variability in recidivism risk due to certain combinations of risk factors. If a youth belongs to a "non-dominant" case configuration (less than 10 observations) the average probability of recidivism for their specific risk score is used as the CACC probability of recidivism in the formula above. As referenced previously, case configurations with less than 10 observations will provide unreliable and variable estimates of the probability of the dependent variable (Miethe, Hart, & Regoeczi, 2008). While not ideal, the average probability for recidivism based on the risk score serves as a more stable measurement.

CACC Risk Scores assume that case configuration probabilities of recidivism found in one sample are able to be generalized to another sample. That means a youth with certain combinations of risk factors in one sample shares the same likelihood of recidivism as a youth who has these same combinations of risk factors in another sample. To test this, conjunctive analysis was run on both the Estimation sample and Validation sample. Case configuration probabilities from the Estimation sample are then compared to the same case configurations from the Validation sample to determine if there is evidence to justify this assumption.

Next, risk scores based on logistic regression are calculated for all youth using the Estimation sample. All seven risk factors from the MJPS are included as independent variables in a single logistic regression model with recidivism as the dependent variable. Next, predicted probabilities are derived from this model for each youth in the dataset. The predicted probabilities of recidivism for each youth obtained from logistic regression serve as the final Logistic Regression Risk Scores.

To transfer CACC, and Logistic Regression Risk Scores from the Estimation to the Validation sample, each case configuration observed in the data was given a unique code. This code is seven digits long and each digit represents the presence (1) or absence (0) of a risk factor in a particular order. For instance, the case configuration that has none of the risk factors present would be coded as 0000000. The case configuration that has all risk factors present would be coded as 1111111. The unique code is used to match the new CACC and logistic regression risk scores calculated on the Estimation sample to those in the Validation sample.

Once all new risk scores are calculated and assigned to each youth in both the Estimation and Validation samples, accuracy comparisons are made. The common statistic measuring risk assessment accuracy is derived from the Receiver Operating Characteristic (ROC) analysis, known as the area under the ROC curve (AUC; see Hanley & McNeil, 1982). The ROC curve is created by plotting the sensitivity (true positive rate) by 1 – specificity (false positive rate) across various cut-points used to classify youth as a recidivist or a non-recidivist (Mossman, 2013). An AUC of 0 indicates perfect negative prediction, 0.5 indicates no better than chance prediction, and an AUC of 1 indicates perfect positive prediction (Van der Put, Van Vugt, Stams, & Van der Laan, 2013). It is common in the criminology and psychology risk assessment literature that an AUC score of 0.7 or above indicates strong prediction performance, between 0.6 and 0.7 indicates moderate performance, and anything below 0.6 indicates poor performance (Barnoski, 2004; Mossman, 2013).

Results

Conjunctive Analysis of Case Configuration

Utilizing the Estimation sample, conjunctive analysis of case configuration (CACC) is run on the seven MJPS risk factors. CACC results are itemized in Table A (appendix C). With seven dichotomously coded independent variables, there are a total of 128 case configurations ($2^7 = 128$) possible. However, not all possible configurations are found in the Estimation sample and several configurations were removed due to small sample sizes. Of the 128 possible configurations, 116 (90.6%) were empirically observed in the data. Of the 116 observed configurations, 62 (53.5%) qualified as “dominant” configurations with the minimum frequency rule of 10 (Miethe et al., 2008). The remaining 54 “non-dominant” case configurations probabilities are replaced with the mean probability of recidivism for the respective MJPS Risk Score.

Table 3.5 displays the number of case configurations (combinations of risk factors) observed that create each of the eight risk scores. Risk scores of 0 and 7 have only one possible configuration each; no risk factors or all risk factors respectively. Risk scores between 1 and 6, however, have multiple ways in which they are obtained, illustrating the limitation of the Burgess Method. While all youth that score a 4 on the risk assessment are predicted to have the same risk of recidivism, there are 14 observed ways in which a score of 4 was reached based on different combinations of risk factors. Similarly, several combinations are empirically observed for risk scores of 5, 3 and 2.

Table 3.5: Case Configurations

Burgess Risk Scores	# Case Configurations
7	1
6	4
5	9
4	14
3	13
2	13
1	7
0	1
Total	62

Different risk factor combinations that comprise the same risk score are only a problem if there is significant variation in the probabilities of recidivism within each score. The matrix developed from CACC is used to investigate the variation in recidivism based on different risk factor combinations. The 62 case configuration matrix is presented in Table A (Appendix C) with configurations sorted by Burgess Risk Score in descending order. It is apparent from this matrix that overall, higher risk scores have higher probabilities of recidivism and lower risk scores have lower probabilities of recidivism. This pattern is expected from an accurate risk assessment. However, certain risk scores have large variations in the probability of recidivism. With a risk score of 4, for example, the 14 risk factor combinations in which a score of 4 can be reached all show variation in the probability of recidivism ranging from a low of 0.23 to a high of 0.73. Each risk score that has more than one case configuration possible shows variation in the probability of recidivism. To visualize the variation in probabilities of recidivism, Figure 3.1 presents a boxplot with each dominant case configuration's probability of recidivism (y axis) grouped by risk level (x axis). The dashed reference line is set at the average probability of recidivism (0.38) for the Estimation sample.

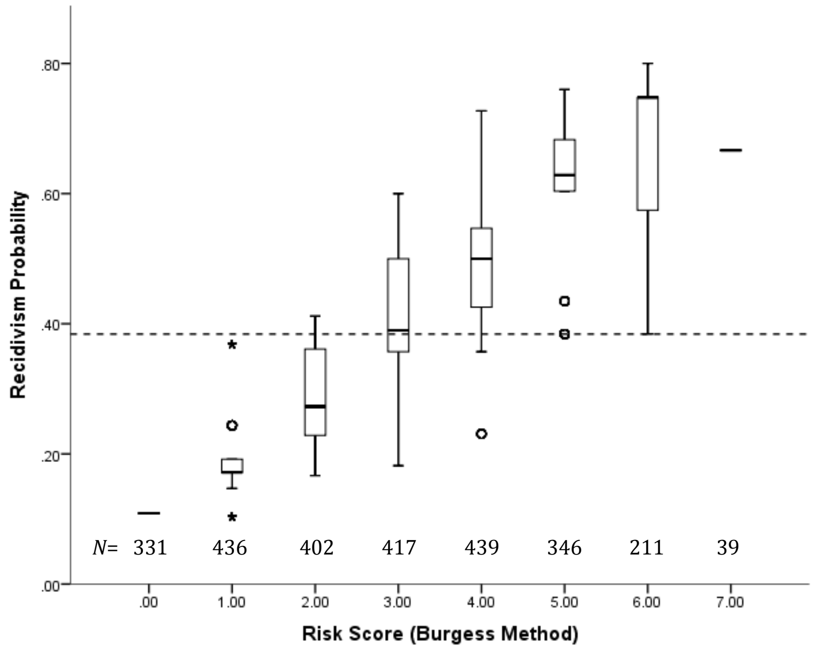

Figure 3.1 Variation in recidivism probabilities within Burgess Risk Scores.

Figure 3.1 displays the extent of variation in recidivism probabilities for each risk score. This figure clearly demonstrates the limitations of the Burgess Risk Score. Not all youth that receive the same risk score have the same likelihood of recidivating. It is important to note that the pattern presented in Figure 3.1 displays an overall increase in recidivism probability as risk scores increase, which is indicative of an accurate assessment. However, the overlap and high variation in probabilities of recidivism among risk scores is evidence that there is a need for improvement.

CACC Risk Score Assumption

The strategy of weighting risk scores based on CACC assumes that case configurations from one sample will have similar probabilities of recidivism as the same case configurations of another sample. For example, the case configurations in the Estimation sample with no risk factors present (0000000) is assumed to have similar probabilities of recidivism when compared to the same case configuration (0000000) in the Validation sample, and this would hold true for all case configurations.

To examine this assumption, bivariate correlation was run on CACC probabilities calculated from the Estimation sample and CACC probabilities calculated from the Validation sample. The correlation between probabilities is 0.924 ($p \leq 0.001$) demonstrating a strong relationship among case configurations. This provides evidence to support the assumption that case configurations have similar probabilities of recidivism and can be generalized from one sample to another.

CACC Risk Score

CACC Risk Scores are calculated for all youth in the Estimation sample (see Analytic Strategy section for equation). The distribution of new risk scores range from 0 to 4.8. To make comparisons between the new CACC Risk Scores and the Burgess Risk Scores, all CACC Risk Scores are recoded into eight categories (0 through 7). Each category is matched in sample size to the previously examined Burgess Risk Scores (Figure 3.1). Figure 3.2 presents the boxplot of case configuration probabilities grouped into the eight CACC risk levels. Similar to Figure 3.1, the dashed reference line presents the average recidivism probability (0.38).

While similar in appearance to Figure 3.1, Figure 3.2 displays less variation in recidivism probability across all eight risk scores. Additionally, there appears to be significantly less overlap among risk score probabilities. It appears that CACC risk scores have partially corrected the issue associated with Burgess Risk Scores in the Estimation sample. CACC scores from the Estimation sample are then merged into the Validation sample to further analyze the accuracy of the new CACC Risk Scores.

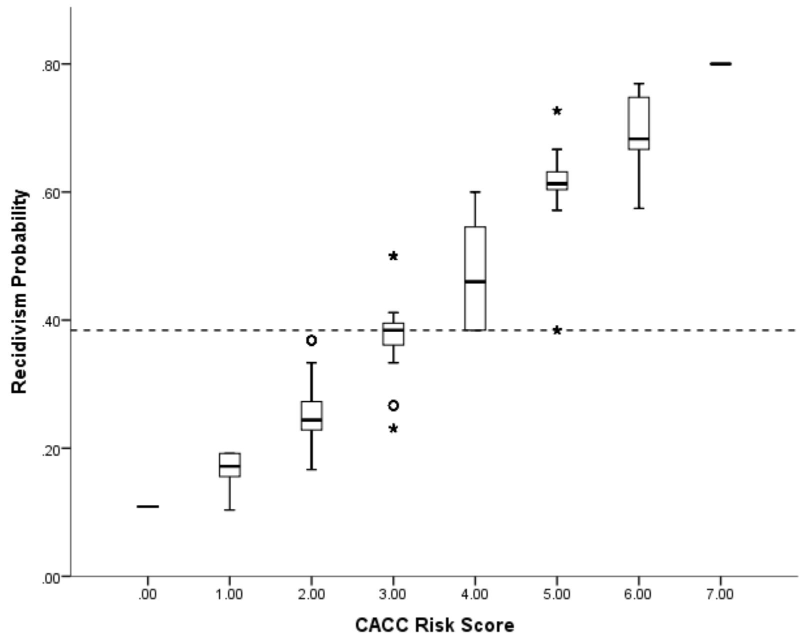

Figure 3.2. Variation in recidivism probabilities within CACC Risk Scores.

Logistic Regression Risk Score

Logistic regression risks scores are created for all youth in the Estimation sample (see Analytic Strategy section for method). All logistic regression risk scores range between a low of 0.12 and a high of 0.79.[22] After the development of risk scores in the Estimation sample, logistic risk scores were transferred to the Validation sample.

Accuracy Comparison

Table 3.6 presents the results from ROC analysis for the Burgess Method, Logistic Regression, and CACC Risk Scores for both the Estimation and Validation samples. Risk scores developed from CACC demonstrate improvement compared to risk scores created from logistic regression and the original Burgess Risk Scores for both Estimation and Validation samples. The highest AUC overall is found for CACC risk scores in the Estimation sample with a 0.748. There is a slight decrease when

[22] A boxplot of Logistic Regression Risk Scores was created similar to the Burgess (Figure 3.1) and CACC Risk Scores (Figure 3.2). However, the boxplot of Logistic Regression Risk scores did not demonstrate any improvement in reducing the variability in recidivism probability from the Burgess Risk scores.

CACC Risk Scores are tested on the Validation sample of 0.744. However, CACC risk scores on the full Validation sample outperform Burgess Risk Scores on the Estimation sample (0.735).

Table 3.6: Risk Score Accuracy Comparison (AUCs)

Sample		MJPS (Burgess Method)	Logistic Risk Score	CACC Risk Score
Estimation				
Full Sample	(*N*=2,621)	.735	.738	.748
Validation				
Full Sample	(*N*=500)	.732	.734	.744
Male	(*n*=301)	.734	.734	.744
Female	(*n*=199)	.727	.734	.743
White	(*n*=396)	.731	.726	.737
Non-White	(*n*=104)	.691	.719	.737

CACC Risk Scores also present improved AUCs for all subsamples in the Validation sample relative to Logistic and Burgess Risk Scores. Logistic Risk Scores are comparable to Burgess Risk Scores showing slight improvement for the full Estimation and Validation samples, in addition to those from the female and non-white samples. Logistic Risk Scores for the male and white samples are either tied to the Burgess Risk Score or show a slight decrease in AUCs. Overall, CACC Risk Score is the preferred method when examining AUCs alone. It is important to note that no sample demonstrates a statistically significant difference among AUCs for any weighting method.[23]

Discussion

This study sought to improve the Burgess Method of weighting risk factors by accounting for the impact that risk factor combinations have on recidivism using conjunctive analysis of case configuration. Findings suggest that relative to the Burgess Method and Logistic Regression method, CACC Risk Scores demonstrated slight improvements in accuracy. While only slight increases in accuracy were noted, the movement marked progress toward a more precise recidivism predictor. This is especially true relative to other studies showing statistical shrinkage in accuracy associated with more complex weighting methods (see Grann & Langstrom, 2007). This study also unveiled several important insights that will serve to further future risk factor weighting and analysis.

[23] Utilizing a difference in areas between two ROC curves calculation (see Hanley & McNeil, 1982).

Insights uncovered by this inquiry include the need to improve the Burgess Method. There are 62 unique observed combinations of risk factors that create the eight MJPS risk scores. This analysis shows there is extreme variation in probabilities of recidivism based on risk factor combinations found on the screener. This fact contrasts the implicit assumption associated with risk assessments, that all youth with the same risk score should be treated the same with respect to their recidivism risk.

Based on the finding of high variation of recidivism within all risk scores on the MJPS, it is the recommended that conjunctive analysis of case configuration become a basic step in the validation of all risk assessments. Beyond the conjunctive matrix provided with the analysis (appendix), this study demonstrates a strategy to visually display recidivism variation probabilities within the same risk score using a boxplot displaying CACC results. An accurate assessment would have the following qualities when examining CACC results through a boxplot (see Figure 3.1 and 3.2): (1) recidivism probability increases as risk scores increase, (2) low variation in recidivism for each case configuration within each risk score, (3) little overlap in probabilities of recidivism among risk scores.

Utilization of conjunctive analysis of case configuration along with the common elements of a typical validation (e.g., ROC analysis) would ensure a more rounded perspective of the screener's ability to predict the outcome of interest. It is likely that any screener created with factors weighted with the Burgess Method will have recidivism variation within each risk score and this will further encourage researchers in the future to find meaningful ways to improve factor weights.

The results from this analysis should not discourage future efforts to attempt to improve the Burgess Method through conjunctive analysis. In fact, several pieces of evidence demonstrate progress in prediction accuracy using this strategy. First, it is shown that there is a very strong correlation ($r > 0.9$) between the probabilities of recidivism from case configurations calculated in one sample to probabilities of recidivism from case configurations calculated from an entirely independent sample. This provides evidence that certain risk factor combinations are inherently more or less predictive of recidivism and the effect of these combinations needs to be accounted for in risk assessments. Next, the effect of multiplying the Burgess Score by the CACC probability appears to reduce the variability in recidivism probability within each risk score as intended (Figure 3.1 and 3.2). The reduction in recidivism variation within risk scores, however, did not significantly improve accuracy results based on ROC analysis.

There are three potential issues that led to the inability to significantly improve risk scores developed from the Burgess Method: (1) The sample size is too small to collect accurate probabilities, (2) Alternate methods of incorporating CACC probabilities would be more effective, and/or (3) No weighting technique will significantly outperform simple methods such as the 0 and 1 scoring system used in the Burgess Method. First, CACC requires a large sample size especially when several independent variables are used, as was the case with this analysis. Seven independent variables create 128 possible case configurations. The Estimation sample for this analysis had 2,621 youth. If all case configurations were observed and all had the same sample size, each configuration would only yield 20 youth. Even in this best-case scenario, 20 youth per configuration would be a small sample size to estimate population case configuration probabilities. As presented in the results, certain configurations were observed at a much higher frequency than others, with 12 configurations not observed at all and 54 configurations with less than 10 youth. If each case configuration had a large sample size to gain an accurate probability of recidivism, CACC Risk Scores may have performed better than found in this analysis with the limited sample size. Next, the evidence presented from CACC demonstrates a need to include an adjustment to the Burgess Risk Score that accounts for the differing probabilities of recidivism based on different configurations. The formula for creating CACC Risk Scores simply takes the probability of recidivism based on the case configuration and multiplies it by the original Burgess Risk Score. This is only one strategy, and alternative methods for including the probability from CACC into the risk score may be beneficial for predictive accuracy. Finally, as has been noted in prior research, advanced techniques for weighting risk factors do not significantly outperform simple techniques such as the Burgess or other non-weighted methods (e.g., Gottfredson & Snyder, 2005; Grann & Langstom, 2008). It may be that prediction accuracy will only be significantly improved upon with more predictive risk factors or by focusing on other decision points in the creation of an assessment.

Limitations and Future Research

Sample size is the primary limitation in this investigation. As discussed above, the small sample size may be problematic for the calculation of accurate probabilities of recidivism for certain case configurations. Additionally, case configurations that did not have at least 10 observations were removed altogether. This left 226 (8.6%) youth in the Estimation sample to use the average probability of recidivism from their respective risk scores instead of a calculated probability from their case configurations. This limitation may have impacted the accuracy of the final CACC Risk Score. Future research attempting a similar weighting strategy should call upon a significantly larger sample size to determine if CACC Risk Scores do in fact increase prediction accuracy compared to simple scores developed from the Burgess Method.

This study was based on the assumption that case configuration probabilities of recidivism were similar across all demographics. However, it is likely that case configurations uniquely affect recidivism between males and females and white and non-white youth respectively. It may be beneficial to address this limitation in future inquiries by examining how demographics affect case configuration probabilities of recidivism and if accuracy can be improved by making demographic adjustments to CACC Risk Scores.

Finally, future research should continue to explore and experiment with alternative methods of weighting risk factors. Risk assessments are being developed and deployed across all levels of the justice system. It is to everyone's benefit to discover and utilize most accurate methods possible.

Conclusion

In an attempt to increase prediction accuracy for youth recidivism, conjunctive analysis of case configuration (CACC) was used to weight risk scores on the Montana Juvenile Probation Screener

(MJPS). The MJPS was originally weighted using the Burgess Method. While the MJPS was found to be an accurate predictor of recidivism (Chapter 3), the Burgess Method does not encapsulate the combinations of risk factors that create each risk score. To address this problem, CACC was used for the specific purpose of adjusting the risk score to account for combinations of risk factors that increase or decrease the likelihood of recidivism. Burgess Risk Scores, along with risk scores derived from logistic regression, were used as a comparison to the risk scores created from conjunctive analysis.

Inquiry results yielded only slight evidence to support the use of CACC as a weighting strategy over the Burgess and Logistic Regression methods. These results may be heavily impacted by the sample size used in this investigation and researchers in the future should attempt his weighting strategy with a significantly larger sample size. This study also demonstrated a new technique to analyze the accuracy of a risk assessment through the use of CACC. Plotting the probabilities of recidivism for each case configuration clustered by risk score in a boxplot provides an excellent visualization of the variation in recidivism probability inherent to each risk score.

Conjunctive analysis of case configuration is still a relatively new strategy, and the utility and potential of it has not been exhausted. This study sought to push the boundaries of CACC by incorporating it in a risk assessment instrument. The goal was to inspire new ways to weight risk factors and analyze accuracy which future researcher can continue to explore.

Appendix C

Table A: Conjuctive Analysis of Case Configuration Matrix

Problem Solve	Anti-Social Friend	More than 1 Misd.	Believes in Fighting	Run Away Kicked Out	First Offense < 13	No Positive Adult Relationships	Risk Score	Prob. Recidivism	*n*
1	1	1	1	1	1	1	7.00	0.67	39
1	1	1	1	0	1	1	6.00	0.80	30
1	1	1	0	1	1	1	6.00	0.77	13
1	1	1	1	1	1	0	6.00	0.75	111
1	1	1	1	1	0	1	6.00	0.57	47
1	1	1	0	1	0	1	5.00	0.76	25
1	1	1	1	1	0	0	5.00	0.68	82
1	0	1	1	1	0	1	5.00	0.67	18
1	1	1	1	0	0	1	5.00	0.65	26
1	1	1	0	0	1	1	5.00	0.63	19
1	1	1	1	0	1	0	5.00	0.63	35
1	0	1	1	1	1	0	5.00	0.63	16
1	1	1	0	1	1	0	5.00	0.60	53
1	1	0	1	1	0	1	5.00	0.43	23
1	0	1	0	0	1	1	4.00	0.73	11
0	1	1	1	1	0	0	4.00	0.67	12
1	0	1	1	0	1	0	4.00	0.61	31
0	1	1	0	1	0	1	4.00	0.60	10
1	1	0	1	0	0	1	4.00	0.59	17
1	1	0	0	1	0	1	4.00	0.57	21
1	1	1	0	1	0	0	4.00	0.55	75
1	1	1	0	0	1	0	4.00	0.55	33
1	1	0	1	0	1	0	4.00	0.50	10
1	1	1	1	0	0	0	4.00	0.46	50
1	1	1	0	0	0	1	4.00	0.43	21
1	1	0	1	1	0	0	4.00	0.43	47
1	0	1	1	1	0	0	4.00	0.36	14
1	0	1	1	0	0	1	4.00	0.23	13
1	0	1	0	0	1	0	3.00	0.60	20
0	1	1	0	1	0	0	3.00	0.54	37
0	1	1	0	0	1	0	3.00	0.53	17
1	1	0	0	1	0	0	3.00	0.50	38
1	0	0	0	0	1	1	3.00	0.50	10
1	1	0	1	0	0	0	3.00	0.40	43
1	1	1	0	0	0	0	3.00	0.39	59
1	0	1	1	0	0	0	3.00	0.36	14
1	1	0	0	0	1	0	3.00	0.33	21
1	0	0	1	1	0	0	3.00	0.27	15
1	1	0	0	0	0	1	3.00	0.24	38
0	1	0	1	1	0	0	3.00	0.20	15
1	0	1	0	1	0	0	3.00	0.18	11
1	0	0	0	1	0	0	2.00	0.41	17
0	1	0	1	0	0	0	2.00	0.38	16
0	1	1	0	0	0	0	2.00	0.36	36
1	1	0	0	0	0	0	2.00	0.33	81
0	0	1	0	1	0	0	2.00	0.27	11
1	0	0	1	0	0	0	2.00	0.27	30
0	1	0	0	1	0	0	2.00	0.26	38
1	0	0	0	0	1	0	2.00	0.26	23
1	0	1	0	0	0	0	2.00	0.23	35
0	0	0	0	0	1	1	2.00	0.20	38
0	1	0	0	0	1	0	2.00	0.20	10
1	0	0	0	0	0	1	2.00	0.18	17
0	1	0	0	0	0	1	2.00	0.17	36
0	0	0	1	0	0	0	1.00	0.37	19
0	0	1	0	0	0	0	1.00	0.24	41
1	0	0	0	0	0	0	1.00	0.19	99
0	1	0	0	0	0	0	1.00	0.17	169
0	0	0	0	0	0	1	1.00	0.16	45
0	0	0	0	1	0	0	1.00	0.15	34
0	0	0	0	0	1	0	1.00	0.10	29
0	0	0	0	0	0	0	0.00	0.11	331

Risk Factors is Present	Risk Factor is Absent

Chapter 5

Conclusion

The juvenile justice system is a maze of decision points beginning at the time of arrest. The decision to detain at the point of arrest and each subsequent decision impacts individual youth and also their communities. Individual decisions may appear logical and to demonstrate unbiased thought and consideration for the best outcome; however, when data are aggregated, patterns of disparities for certain youth often emerge. It's unlikely these disparities arise out of malicious intent from officers in the field. It's more likely that these disparities are an artifact of an unstandardized, subjective decision process that is easily influenced by extralegal factors. Three separate studies lay a roadmap for a shift to evidence-based decision making in the juvenile justice system, by: (1) demonstrating the need for decisions to be improved, (2) incorporating data-driven strategies to improve decisions, and (3) continuously developing new ways to improve future data-driven strategies.

Chapter 2 begins with an investigation on the decision to initially detain youth in Montana. This study adds to the body of knowledge on initial detention by utilizing a population of youth from a smaller rural state where American Indians are the largest minority subsample. Eight years of initial detention decisions (N = 26,128) are included in this analysis. The findings indicate that the severity of current offense and the existence of prior offenses constitute the greatest predictors of initial detention. However, results from this inquiry also provide strong evidence to support the hypothesis that proximately to detention facility influences detention decisions. The closer an arrest is to a detention facility, the more likely a youth will be placed in initial detention, even when controlling for offense severity and prior offenses. Previous studies have demonstrated a divide among rural, urban, and suburban populations and justice system decisions. This study finds that when controlling for distance to facility, being arrested in an urban (metro) county no longer affects the likelihood of detention relative to being arrested in a rural (non-core) county. In agreement with literature on Disproportionate Minority Contact (DMC), the results demonstrate a disproportionate amount of non-white youth are detained in Montana. Interestingly, race moderated the effect that distance to detention has on detention decisions. There is greater disparity in initial detention between white and non-white youth in areas that are closer to a detention facility than those areas further away. Overall, the results demonstrate the complexity of this initial decision in the juvenile justice system. While factors such as current offense and offense history are being used to make the decision to detain a youth, geography and race are also shown to influence these decisions. Future work must be done to determine why such disparities are occurring and what can be done to ensure they do not continue.

Chapter 3 continues by developing a short recidivism risk screener, the Montana Juvenile Probation Screener (MJPS), to assist juvenile probation officers with placement decisions (e.g., formal or informal probation). This study utilizes six years of juvenile intakes ($N = 3{,}121$) to create and validate a screener that can be administered quickly to youth to determine recidivism risk. Risk factors included in the new screeners were selected through a series of bivariate and multivariate models. Out of the 246 risk factors eligible for the screener, seven factors were found to be important predictors at both the bivariate and multivariate level. Through a series of validation measures, the MJPS demonstrated increased prediction accuracy in the Estimation and Validation samples when compared to the current risk assessment (pre-screen BOT), which requires probation officers to locate information on approximately 40 risk factors. A recommendation stemming from this study is that the MJPS be piloted to better understand its abilities in a prospective validation study. A demonstration of validity in the field may increase fidelity of use among juvenile probation officers in Montana, which could assist in a more informed structured decision process.

Finally, Chapter 4 analyzes an experimental method of weighting risk factors using conjunctive analysis of case configuration (CACC). This study seeks to increase risk prediction accuracy from the original Burgess Method Risk Score on the MJPS created in Chapter 3. The Burgess Method utilizes a simple 0, 1 scoring method for all risk assessment factors. One limitations of this method is that different risk factor combinations can be placed together to calculate the same risk score as another combination of factors. This analysis demonstrated that these differing combinations affected the likelihood of recidivism. CACC was used to address this limitation by adjusting the risk score to account for different combinations shown to increase or decrease risk of recidivism. Using six years of juvenile intakes (N = 3,121), new risk scores based on CACC were created. Those scores were then compared against the original Burgess Method Risk Score and also a risk score derived from logistic regression. CACC Risk Scores demonstrate significantly less variation in the probability of recidivism within risk scores but only minor accuracy improvements over the Burgess Method and Logistic Regression Risk Scores. While accuracy improvement was not significantly increased, this study demonstrates that combinations of factors are important to consider in any risk assessment. Furthermore, this analysis provided a new strategy for analyzing and validating a risk assessment using CACC. This research is limited in its sample size, and future research using a similar weighting strategy should be employed with a significantly larger sample size.

Overall, this inquiry demonstrates how much work there is left to do. Researchers must continue to investigate how decisions are made at every point of the justice system and strategies for making such decisions increasingly objective and accurate. A more informed and objective decision-making process will not completely alleviate the over incarceration problem in the United States where seven million people are under justice system supervision (Glaze & Kaeble, 2014). However, improving justice system decision making could increase objectivity and decrease disparities; reduce unnecessary supervision, while keeping the community safe, and assist in making an increasingly fair and just system.

References

Aizer, A., & Doyle, J. (2013). What is the long-term impact of incarcerating juveniles. *Retrieved June*, *15*.

Andrews, D. A., Bonta, J., & Hoge, R. D. (1990). Classification for effective rehabilitation: Rediscovering psychology. *Criminal justice and Behavior*, *17*(1), 19-52.

Andrews, D. A., Bonta, J., & Wormith, J. S. (2004). Level of service/case management inventory: An offender assessment system. *North Tonawanda, NY: Multi-Health Systems*.

Andrews, D. A., Bonta, J., & Wormith, J. S. (2006). The recent past and near future of risk and/or need assessment. *Crime & Delinquency*, *52*(1), 7-27.

Annie E. Casey Foundation (2013). *Reducing youth incarceration in the United States.* Retrieved from https://www.aecf.org/resources/reducing-youth-incarceration-in-the-united-states/

Armstrong, G. S., and Rodriguez, N. (2005) Effects of individual and contextual characteristics on preadjudication detention of juvenile delinquents. *Justice Quarterly,* 22(4), 521-539.

Ayres, I., & Borowsky, J. (2008). A study of racially disparate outcomes in the Los Angeles Police Department. Los Angeles, CA: ACLU of Southern California.

Baird C. (2018). The CAIS/JAIS Approach to Assessment. *Handbook of Recidivism Risk/Needs Assessment Tools*, 33.

Barnes, G., & Hyatt, J. (2012). Using random forest risk prediction in the Philadelphia probation department. *Washington, DC: National Criminal Justice Reference Service, Document NCJ241346*.

Barnoski, R. (2004). Assessing risk for re-offense: Validating the Washington State Juvenile Court assessment appendices. *Olympia, WA: Washington State Institute for Public Policy*.

Barnoski, R. P. (2009). *Providing evidence-based programs with fidelity in Washington State juvenile courts: Cost analysis*. Olympia: Washington State Institute for Public Policy.

Barnoski, R. P., & Drake, E. K. (2007). *Washington's Offender Accountability Act: Department of Corrections' Static Risk Instrument*. Washington State Institute for Public Policy.

Barnoski, R., & Aos, S. (2004). Outcome evaluation of Washington State's research-based programs for juvenile offenders. *Olympia, WA: Washington State Institute for Public Policy*, 460.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology*, 51(6), 1173.

Berk, R., Sherman, L., Barnes, G., Kurtz, E., & Ahlman, L. (2009). Forecasting murder within a population of probationers and parolees: a high stakes application of statistical learning. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *172*(1), 191-211.

Bezruki, D., Varana, D., & Hill, C. (1999). An evaluation of secure juvenile detention. Madison WI: Legislative Audit Bureau.

Bigbee, J. L., Gehrke, P., & Otterness, N. (2009). Public health nurses in rural/frontier one-nurse offices. *Rural and Remote Health*, *9*(1282), 1-11.

Bishop, D. M. (2005). The role of race and ethnicity in juvenile justice processing. Our children, their children: Confronting racial and ethnic differences in American juvenile justice, 23-82.

Bishop, D. M., & Frazier, C. E. (1988). The influence of race in juvenile justice processing. *Journal of Research in Crime and Delinquency*, 25(3), 242-263.

Bishop, D. M., & Frazier, C. E. (1995). Race effects in juvenile justice decision-making: Findings of a statewide analysis. *J. Crim. l. & Criminology*, 86(2), 392-414.

Bishop, M. A., & Trout, J. D. (2002). 50 years of successful predictive modeling should be enough: Lessons for philosophy of science. *Philosophy of Science*, *69*(S3), S197-S208.

Blackmon, B. J., Robison, S. B., & Rhodes, J. L. (2016). Examining the influence of risk factors across rural and urban communities. *Journal of the Society for Social Work and Research*, 7(4), 615-638.

Bortner, M. A., & Wornie, L. R. (1985). The preeminence of process: An example of refocused justice research. *Social Science Quarterly*, 66(2), 413.

Burgess, E. W. (1928). Factors determining success or failure on parole. *The workings of the indeterminate sentence law and the parole system in Illinois*, 221-234.

Campbell, N. A., Barnes, A. R., Mandalari, A., Onifade, E., Campbell, C. A., Anderson, V. R., ... & Davidson, W. S. (2018). Disproportionate minority contact in the juvenile justice system: An investigation of ethnic disparity in program referral at disposition. *Journal of Ethnicity in Criminal Justice*, 16(2), 77-98.

Caulkins, J., Cohen, J., Gorr, W., & Wei, J. (1996). Predicting criminal recidivism: A comparison of neural network models with statistical methods. *Journal of Criminal Justice*, *24*(3), 227-240.

Chesney-Lind, M. (1977). Judicial paternalism and the female status offender: Training women to know their place. *Crime & Delinquency*, 23(2), 121-130.

Chung, H. L., Little, M., & Steinberg, L. (2005). The transition to adulthood for adolescents in the juvenile justice system: A developmental perspective. On your own without a net: The transition to adulthood for vulnerable populations, 68-91.

Cocozza, J. J., & Shufelt, J. L. (2006). Juvenile mental health courts: An emerging strategy. Delmar, NY: National Center for Mental Health and Juvenile Justice.

Cocozza, J. J., & Skowyra, K. R. (2000). Youth with mental health disorders: Issues and emerging responses. *Juvenile Justice*, 7, 3.

Cohen, L. E., & Felson, M. (1979). Social change and crime rate trends: A routine activity approach. *American sociological review*, 588-608.

Cohen, T. H., Lowenkamp, C. T., & Robinson, C. (2018). The Federal Post-Conviction Risk Assessment Instrument: A Tool for Predicting Recidivism for Offenders on Federal Supervision. *Handbook of Recidivism Risk/Needs Assessment Tools*, 77.

Cromartie, J., Nulph, D., Hart, G., & Dobis, E. (2013). Defining frontier areas in the United States. *Journal of Maps*, 9(2), 149-153.

Dannefer, D., & Schutt, R. K. (1982). Race and juvenile justice processing in court and police agencies. *American Journal of Sociology*, 87(5), 1113-1132.

Davis, J., & Sorensen, J. R. (2013). Disproportionate minority confinement of juveniles: A national examination of Black–White disparity in placements, 1997-2006. *Crime & Delinquency,* 59(1), 115-139.

Dawes, M., Robyn. (1979). The Robust Beauty of Improper Linear Models in Decision Making. *American Psychologist,* 34(7), 571-582.

DeJong, C., & Jackson, K. C. (1998). Putting race into context: Race, juvenile justice processing, and urbanization. *Justice Quarterly*, 15(3), 487-504.

Development Services Group, Inc. (2014). Disproportionate Minority Contact. Washington, D.C.: Office of Juvenile Justice and Delinquency Prevention. https://www.ojjdp.gov/mpg/litreviews/Disproportionate_Minority_Contact.pdfPrepared by Development Services Group, Inc., under cooperative agreement number 2013–JF–FX–K002.

Dishion, T. J., McCord, J., & Poulin, F. (1999). When interventions harm: Peer groups and problem behavior. *American psychologist*, 54(9), 755.

Douglas, K. S., Reeves, K., & Otto, R. (2010). Historical-clinical-risk management-20 (HCR-20) violence risk assessment scheme. *Rationale, Application, and Empirical Overview, eds RK Otto and KS Douglas (New York, NY: Taylor & Francis)*, 147-186.

Drawve, G., Thomas, S. A., & Hart, T. C. (2017). Routine activity theory and the likelihood of arrest: A replication and extension with conjunctive methods. *Journal of contemporary criminal justice*, *33*(2), 121-132.

Drawve, G., Thomas, S. A., & Walker, J. T. (2014). The likelihood of arrest: A routine activity theory approach. *American Journal of Criminal Justice*, *39*(3), 450-470.

Duwe, G., & Freske, P. J. (2012). Using logistic regression modeling to predict sexual recidivism: The Minnesota sex offender screening tool-3 (MnSOST-3). *Sexual Abuse*, *24*(4), 350-377.

Duwe, G., & Freske, P. J. (2012). Using logistic regression modeling to predict sexual recidivism: The Minnesota sex offender screening tool-3 (MnSOST-3). *Sexual Abuse*, *24*(4), 350-377.

Duwe, G., & Kim, K. (2016). Sacrificing accuracy for transparency in recidivism risk assessment: The impact of classification method on predictive performance. *Corrections*, *1*(3), 155-176.

Feld, B. C. (1991). Justice by geography: Urban, suburban, and rural variations in juvenile justice administration. *J. Crim. L. & Criminology*, 82, 156.

Fite, P. J., Wynn, P., & Pardini, D. A. (2009). Explaining discrepancies in arrest rates between Black and White male juveniles. *Journal of consulting and clinical psychology*, 77(5), 916.

Forrest, C. B., Tambor, E., Riley, A. W., Ensminger, M. E., & Starfield, B. (2000). The health profile of incarcerated male youths. *Pediatrics,* 105(1), 286-291.

Frazier, C. E., & Bishop, D. M. (1985). Pretrial Detention of Juveniles and Its Impact on Case Dispositions, *The. J. Crim. L. & Criminology*, 76, 1132.

Friedman, R. M., Katz-Leavy, J. W., Manderscheid, R. W., & Sondheimer, D. L. (1996). Prevalence of serious emotional disturbance in children and adolescents. Mental Health, United States, 996, 71-89.

Gaarder, E., Rodriguez, N., & Zatz, M. S. (2004). Criers, liars, and manipulators: Probation officers' views of girls. *Justice Quarterly*, 21(3), 547-578.

Gatti, U., Tremblay, R. E., & Vitaro, F. (2009). Iatrogenic effect of juvenile justice. *Journal of Child Psychology and Psychiatry*, 50(8), 991-998.

Glaze, L. E. & Kaeble, D. (2014). *Correctional population in the United States, 2013* (NCJ 248479). Washington, DC: Bureau of Justice Statistics.

Golub, A. (1990). The termination rate of adult criminal careers. School of Urban and Public Affairs, Carnegie Mellon University.

Gordon R. (2016). Utah Juvenile Justice Working Group: Final Report. Utah Juvenile Justice Working Group and the Pew Charitable Trusts and the Crime and Justice Institute at CRJ.

Gottfredson, D. M., & Snyder, H. N. (2005). The Mathematics of Risk Classification: Changing Data into Valid Instruments for Juvenile Courts. NCJ 209158. *Office of Juvenile Justice and Delinquency Prevention*.

Gottfredson, S. D., & Gottfredson, D. M. (1980). Screening for risk: A comparison of methods. *Criminal Justice and Behavior*, *7*(3), 315-330.

Grann, M., & Långström, N. (2007). Actuarial assessment of violence risk: To weigh or not to weigh?. *Criminal Justice and Behavior*, *34*(1), 22-36.

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, *143*(1), 29-36.

Hart, T. C., Rennison, C. M., & Miethe, T. D. (2017). Identifying patterns of situational clustering and contextual variability in criminological data: An overview of conjunctive analysis of case configurations. *Journal of contemporary criminal justice*, *33*(2), 112-120.

Hilton, N. Z., Harris, G. T., & Rice, M. E. (2006). Sixty-six years of research on the clinical versus actuarial prediction of violence. *The Counseling Psychologist*, *34*(3), 400-409.

Hockenberry, S., & Puzzanchera, C. (2015). Juvenile court statistics 2013.

Hollist, D., Coolidge, J., Delano, W., Greenwood, I., King, M., McLean, T., McKay, P., Burfeind, J., Harris, C., & Doyle, D. (2012). "Assessing the Mechanisms That Contribute to Disproportionate Minority Contact in Montana's Juvenile Justice Systems." Retrieved from: http://www.mbcc.mt.gov/Data/SAC/RAI/DMCAssessRep.pdf

Holman, B., & Zeidenberg, J. (2006). The dangers of detention: The impact of incarcerating youth in detention and other secure facilities. Justice Policy Institute.

Howard, P. D., & Dixon, L. (2012). The construction and validation of the OASys Violence Predictor: Advancing violence risk assessment in the English and Welsh correctional services. *Criminal Justice and Behavior*, *39*(3), 287-307.

Huizinga, D., Thornberry, T., Knight, K., Lovegrove, P., Loeber, R., Hill, K., & Farrington, D. P. (2007). Disproportionate minority contact in the juvenile justice system: A study of differential minority arrest/referral to court in three cities: A report to the office of juvenile justice and delinquency prevention. US Department of Justice, NCJRS.

Johnson, J. B., & Secret, P. E. (1995). The effects of court structure on juvenile court decision making. *Journal of Criminal Justice*, 23(1), 63-82.

Kakade, M., Duarte, C. S., Liu, X., Fuller, C. J., Drucker, E., Hoven, C. W., ... & Wu, P. (2012). Adolescent substance use and other illegal behaviors and racial disparities in criminal justice system involvement: Findings from a US national survey. *American Journal of Public Health*, 102(7), 1307-1310.

Kaufmann, Charles F. (1997). Disproportionate Minority Confinement in Massachusetts. Boston, Mass: Statistical Analysis Center.

Kempf-Leonard, K. (2007). Minority youths and juvenile justice: Disproportionate minority contact after nearly 20 years of reform efforts. *Youth Violence and Juvenile Justice,* 5(1), 71-87.

King, E., Shaler, G., & Dumont, R. (2015). Disproportionate Contact: Youth of Color in Maine's Juvenile Justice System.

Lambie, I., & Randell, I. (2013). The impact of incarceration on juvenile offenders. *Clinical Psychology Review*, *33*(3), 448-459.

Latessa, E. J., Lovins, B., & Lux, J. (2018). The Ohio Risk Assessment System. *Handbook of Recidivism Risk/Needs Assessment Tools*, 147.

Leiber, M. J. (2002). Disproportionate minority confinement (DMC) of youth: An analysis of state and federal efforts to address the issue. *Crime & Delinquency*, 48(1), 3-45.

Leiber, M. J. (2003). The contexts of juvenile justice decision making: When race matters. SUNY Press.

Leiber, M. J., & Boggess, L. N. (2012). Race, probation violations, and structured secure detention decision making in three jurisdictions. *Youth Violence and Juvenile Justice*, 10(4), 333-353.

Leiber, M. J., & Fox, K. C. (2005). Race and the impact of detention on juvenile justice decision making. *Crime & Delinquency*, 51(4), 470-497.

Leiber, M. J., & Johnson, J. D. (2008). Being young and black: What are their effects on juvenile justice decision making?. *Crime & Delinquency*, 54(4), 560-581.

Leiber, M. J., & Peck, J. H. (2012). Race in juvenile justice and sentencing policy: An overview of research and policy recommendations. *Law & Ineq*., 31, 331.

Leiber, M., Bishop, D., & Chamlin, M. B. (2011). Juvenile justice decision-making before and after the implementation of the disproportionate minority contact (DMC) mandate. *Justice Quarterly*, 28(3), 460-492.

Leonard, K. K. E., Pope, C. E., & Feyerherm, W. H. (1995). Minorities in juvenile justice. Sage Publications, Inc.

Lipsey, M. W. (2009). The primary factors that characterize effective interventions with juvenile offenders: A meta-analytic overview. *Victims and offenders,* 4(2), 124-147.

Lipsey, M. W., & Cullen, F. T. (2007). The effectiveness of correctional rehabilitation: A review of systematic reviews. *Annu. Rev. Law Soc. Sci.,* 3, 297-320.

Lipsey, M. W., & Wilson, D. B. (1998). *Effective intervention for serious juvenile offenders: A synthesis of research*. Sage Publications, Inc.

Loughran, T. A., Mulvey, E. P., Schubert, C. A., Fagan, J., Piquero, A. R., & Losoya, S. H. (2009). Estimating a dose-response relationship between length of stay and future recidivism in serious juvenile offenders. *Criminology*, 47(3), 699-740.

MacDonald, J. M., & Chesney-Lind, M. (2001). Gender bias and juvenile justice revisited: A multiyear analysis. *Crime & Delinquency*, 47(2), 173-195.

MacKenzie, D. L., Wilson, D. B., & Kider, S. B. (2001). Effects of correctional boot camps on offending. *The Annals of the American Academy of Political and Social Science,* 578(1), 126-143.

Mallett, C. A., & Stoddard-Dare, P. (2010). Predicting secure detention placement for African-American juvenile offenders: Addressing the disproportionate minority confinement problem. *Journal of Ethnicity in Criminal Justice*, 8(2), 91-103.

MCA §41-5-102.3

MCA §41-5-322.2

MCA §41-5-332.1

McCarthy, B. R., & Smith, B. L. (1986). The conceptualization of discrimination in the juvenile justice process: The impact of administrative factors and screening decisions on juvenile court dispositions. *Criminology,* 24(1), 41-64.

McGuire, M. D. (2002). Cumulative disadvantage as an explanation for observed disproportionality within the juvenile justice system: An empirical test. *Juvenile and Family Court Journal*, 53(1), 1-17.

McKay, P. D. (2012). Rescoring the Montana Risk Assessment Instrument: A Comparison of Methods. *University of Montana Thesis.*

McKay, P., Hollist, D., Bunch, J., Acton, D., Tillman, T., & Harris, C. (2015). The Montana Back on Track Risk Assessment Instrument: An Assessment and Validation Study.

Mendel, R. A. (2009). Two decades of JDAI: From demonstration project to national standard.

Mendel, R. A. (2011). No place for kids: The case for reducing juvenile incarceration. Annie E. Casey Foundation, 24.

Midi, H., Sarkar, S. K., & Rana, S. (2010). Collinearity diagnostics of binary logistic regression model. *Journal of Interdisciplinary Mathematics,* 13(3), 253-267.

Miethe, T. D., Hart, T. C., & Regoeczi, W. C. (2008). The conjunctive analysis of case configurations: An exploratory method for discrete multivariate analyses of crime data. *Journal of Quantitative Criminology*, *24*(2), 227-241.

Miller, J., & Maloney, C. (2013). Practitioner compliance with risk/needs assessment tools: A theoretical and empirical assessment. *Criminal Justice and Behavior*, *40*(7), 716-736.

Montana State Library (n.d). *Geographic Information Clearing House.* Retrieved from http://geoinfo.msl.mt.gov/.

Moore, L. D., & Padavic, I. (2010). Racial and ethnic disparities in girls' sentencing in the juvenile justice system. *Feminist Criminology*, 5(3), 263-285.

Mossman, D. (2013). Evaluating risk assessments using receiver operating characteristic analysis: Rationale, advantages, insights, and limitations. *Behavioral Sciences & the Law*, *31*(1), 23-39.

Munyo, I. (2015). The juvenile crime dilemma. *Review of Economic Dynamics*, *18*(2), 201-211.

National Research Council. (2013). Reforming juvenile justice: A developmental approach. National Academies Press.

Nuffield, J. (1982). *Parole decision-making in Canada: Research towards decision guidelines*. Ottawa, Canada: Communication Division, Solicitor General of Canada.

Office of Management and Budget (OMB) University of Montana. (2015). *Defining Rural: why do these definitions matter?* Retrieved from http://rtc.ruralinstitute.umt.edu/resources/defining-rural/

Parent, D., Leiter, V., Kennedy, S., Livens, L., Wentworth, D., & Wilcox, S. (1994). Conditions of confinement: Juvenile detention and corrections facilities [Research Report]. Washington, DC: US Department of Justice, Office of Juvenile Justice and Delinquency Prevention.

Pope, C. E., & Feyerherm, W. (1993). Minorities and the Juvenile Justice System Washington. DC: US Department of Justice, Office of Juvenile Justice and Delinquency NCJ, 145849.

Pope, C. E., & Feyerherm, W. H. (1990). Minority status and juvenile justice processing: An assessment of the research literature (Part I). In *Criminal Justice Abstracts* (Vol. 22, No. 2, pp. 327-335).

Puzzanchera, C., & Adams, B. (2009). Juvenile arrests. Juvenile offenders and victims: national report series bulletin. Washington, DC: US Department of Justice, Office of Justice Programs.

Ragin, C. C. (2013). New directions in the logic of social inquiry. *Political Research Quarterly*, 171-174.

Rice, M. E., Harris, G. T., & Lang, C. (2013). Validation of and revision to the VRAG and SORAG: The Violence Risk Appraisal Guide—Revised (VRAG-R). *Psychological assessment*, *25*(3), 951.

Rodriguez, N. (2007). Restorative justice at work: Examining the impact of restorative justice resolutions on juvenile recidivism. *Crime & Delinquency*, 53(3), 355-379.

Rodriguez, N. (2010). The cumulative effect of race and ethnicity in juvenile court outcomes and why preadjudication detention matters. *Journal of research in crime and delinquency*, 47(3), 391-413.

Sampson, R. J., & Lauritsen, J. L. (1997). Racial and ethnic disparities in crime and criminal justice in the United States. *Crime and justice*, 21, 311-374.

Shelden, R. G. (1999). Detention diversion advocacy: An evaluation. Washington, DC: US Department of Justice, Office of Justice Programs, Office of Juvenile Justice and Delinquency Prevention.

Shook, J. J., & Goodkind, S. A. (2009). Racial disproportionality in juvenile justice: The interaction of race and geography in pretrial detention for violent and serious offenses. *Race and social problems*, 1(4), 257.

Sickmund, M. (2009). Delinquency cases in juvenile court, 2005. *Drugs,* 17, 20.

Sickmund, M., Sladky, T. J., Kang, W., & Puzzanchera, C. (2011). Easy access to the census of juveniles in residential placement. Retrieved February, 14, 2018.

Sickmund, M., Snyder, H. N., & Poe-Yamagata, E. (1997). Juvenile Offenders and Victims: 1997 Update on Violence. Statistics Summary.

Silver, E., & Chow-Martin, L. (2002). A multiple models approach to assessing recidivism risk: Implications for judicial decision making. *Criminal justice and behavior*, *29*(5), 538-568.

Silver, E., Smith, W. R., & Banks, S. (2000). Constructing actuarial devices for predicting recidivism: A comparison of methods. *Criminal Justice and behavior*, *27*(6), 733-764.

Simon, F. H. (1972). Statistical methods of making prediction instruments. *Journal of Research in Crime and Delinquency*, *9*(1), 46-53.

Singh, J., P., Kroner, D., G., Wormith, J., S., Desmarais, S., L., & Hamilton, Z. (2018). *Handbook of Recidivism Risk/Needs Assessment Tools.* West Sussex: John Wiley & Sons.

Steadman, H. J., Silver, E., Monahan, J., Appelbaum, P., Robbins, P. C., Mulvey, E. P., ... & Banks, S. (2000). A classification tree approach to the development of actuarial violence risk assessment tools. *Law and human behavior*, *24*(1), 83-100.

Steinhart, D. (2006). Juvenile detention risk assessment: A practice guide to juvenile detention reform. *Juvenile Detention Alternatives Initiative A project of the Annie E Casey Foundation*, *28*, 2011.

Tam, C. C., Abrams, L. S., Freisthler, B., & Ryan, J. P. (2016). Juvenile justice sentencing: Do gender and child welfare involvement matter? *Children and Youth Services Review*, 64, 60-65.

Tandy, K. B. (2013). Do No Harm: The Enhanced Application of Legal and Professional Standards in Protecting Youth from the Harm of Isolation in Youth Correctional Facilities. *Children's Legal Rights Journal*, 34, 143.

Teplin, L. A., Abram, K. M., McClelland, G. M., Dulcan, M. K., & Mericle, A. A. (2002). Psychiatric disorders in youth in juvenile detention. *Archives of general psychiatry*, 59(12), 1133-1143.

United States Census. (2017). American Community Survey Population Estimates. Retrieved January 23rd 2017.

Van Der Put, C. E., Deković, M., Stams, G. J. J., Van Der Laan, P. H., Hoeve, M., & Van Amelsfort, L. (2011). Changes in risk factors during adolescence: Implications for risk assessment. *Criminal justice and behavior*, *38*(3), 248-262.

Van der Put, C. E., Van Vugt, E. S., Stams, G. J. J., Deković, M., & Van der Laan, P. H. (2013). Short-term general recidivism risk of juvenile sex offenders: Validation of the Washington State Juvenile Court Prescreen Assessment. *International Journal of Offender Therapy and Comparative Criminology*, *57*(11), 1374-1392.

VanNostrand, M., & Rose, K. (2009). Pretrial risk assessment in Virginia: Report for the Virginia Department of Criminal Justice Services.

Vieira, T. A., Skilling, T. A., & Peterson-Badali, M. (2009). Matching court-ordered services with treatment needs: Predicting treatment success with young offenders. *Criminal justice and behavior*, 36(4), 385-401.

VonNostrand, M., & Lowenkamp, C. (2013). Assessing pretrial risk without a defendant interview.

Walmsly, R. (2013). *World prison population list* (2nd ed.). London, England: International Centre for Prison Studies.

Wilcox, D., Beech, A., Markall, H. F., & Blacker, J. (2009). Actuarial risk assessment and recidivism in a sample of UK intellectually disabled sexual offenders. *Journal of Sexual Aggression*, *15*(1), 97-106.

Williams, L. R., LeCroy, C. W., & Vivian, J. P. (2014). Assessing risk of recidivism among juvenile offenders: The development and validation of the recidivism risk instrument. *Journal of evidence-based social work*, *11*(4), 318-327.

Wood, J., Foy, D. W., Goguen, C. A., Pynoos, R., & James, C. B. (2002). Violence exposure and PTSD among delinquent girls. *Journal of Aggression, Maltreatment & Trauma,* 6(1), 109-126.

Wordes, M., Bynum, T. S., & Corley, C. J. (1994). Locking up youth: The impact of race on detention decisions. *Journal of Research in Crime and Delinquency,* 31(2), 149-165.

Wormith, J. S., & Bonta, J. (2018). The Level of Service (LS) Instruments. *Handbook of recidivism risk/needs assessment tools*, 117-145.

Wu, B. (1997). The effect of race on juvenile justice processing. *Juvenile and Family Court Journal*, 48(1), 43-52.